Final Report

Peer Review of Michigan State University's PCB Exposure and Effects Studies in the Floodplain of the Kalamazoo River

December 1, 2008

Peer Review Panel

Dr. Ken Dickson, University of North Texas, Peer Review Manager

Dr. Larry Barnthouse, LWB Environmental Services, Inc.

Dr. Anne Fairbrother, Parametrix Inc.

Dr. Keith Grasman, Calvin College

Dr. Mark Harwell, Harwell Gentile & Associates, LC

Dr. Lyman McDonald, West Inc.

Dr. Charles Menzie, Exponent Inc.

Dr. William Warren-Hicks, EcoStat, Inc.

Table of Contents

1.0 Executive Summary

The Peer Review Panel (Panel) recognizes the considerable level of effort committed by the Michigan State University (MSU) team to gather site-specific information in support of assessing risks to terrestrial wildlife in the Kalamazoo River Superfund Site (KRSS). The Panel concludes that most of these data do meet the scientific standards appropriate for incorporation into a weight-of-evidence approach to risk management, including the important addition of an area-specific risk assessment at Trowbridge.

The work by the MSU team provides useful data both for quantitative exposure estimates and for a qualitative, weight-of-evidence approach to estimating effects. The published papers provided to the Panel each contribute some data and information that could be used in a weight-of-evidence assessment of risks along with other data and approaches used in the Baseline Ecological Risk Assessment (BERA). However, there remain large uncertainties in the MSU data, issues with the MSU study designs, and insufficient documentation (and lack of agreement) among Toxicity Reference Value (TRV) derivations. Consequently, the MSU studies are not adequate for use as stand-alone documentation for a quantitative risk analysis, as discussed further below. On the other hand, MSU has not yet taken full advantage of the data that they have developed, and there is the potential for additional analyses that would enhance their contribution in a weight-of-evidence risk assessment.

The Panel identified the following as the primary strengths of the MSU field studies:

1. An important strength of the MSU studies is that they focused on directly measuring aspects of several endpoints of concern in the real-world environment. The field studies provide site-specific direct measurements of PCB concentrations in eggs, soil and prey items. Measurements of the endpoints of concern provide investigators with the ability to extend the statistical and modeling analyses beyond those used by USEPA. In addition, the MSU data provide the analysis of actual diets of receptors of choice from the field studies. The field-collected information incorporates soil-specific effects (e.g., soil carbon), congener-specific differences in accumulation rates, and species-specific information on diets related to the Trowbridge site. Each of these facets of the MSU studies can provide valuable input into the weight-of-evidence risk assessment and decision-making process.

2. MSU's congener-specific data are a major strength that allow the examination of congener patterns, total TEQs, and the contribution of individual congeners to total TEQs at different trophic levels. This congener-specific approach reflects the current state-of-the-science in this field and provides an important supplementary line-of-evidence for evaluating exposure even if Michigan regulations focus on total PCBs.

3. Studies of productivity of the bluebirds, wrens, and great horned owls potentially provide useful, qualitative evidence of reproductive performance of on-site species. The strength

of these studies is that they are directly measuring one of the assessment endpoints ("do PCBs affect reproduction of birds?").

4. Data on concentrations of PCBs in shrews and soils provide an opportunity to explore the development of spatially explicit shrew:soil bioaccumulation functions that might have applicability to other former impoundments. This may also reduce some of the uncertainty associated with bioavailability of PCB congeners in soils.

The following items summarize the primary limitations of the MSU studies:

1. There is an absence of a comprehensive conceptual ecological model identifying pathways of exposures and relating effects to endpoints of concern. The data analysis and risk assessment methods employed by MSU should be consistent with the underlying conceptual model. Unfortunately, without an explicit conceptual model, interpretation of the analytical results found in the MSU papers is difficult, if not compromised.

2. The nest productivity studies are limited by small samples sizes (which in some cases are insufficient to draw defensible conclusions) and significant differences in habitats between the Trowbridge impoundment and reference area. There are issues with pseudo-replication and other aspects of study design as well as the calculation and analysis of reproductive parameters.

3. Reliance on species for which aquatic organisms are a significant portion of the diet (specifically bluebirds, and owls) and lack of direct measurements of PCB concentrations in diets of robins (the most highly exposed avian species addressed in the BERA) limit the utility of these studies in assessing risks associated with terrestrial exposure pathways.

4. If the former impoundments change to a more terrestrial environment over time, as anticipated, it will be important to consider the potential for changes from the current species assemblage. Therefore, while the species utilized in MSU give a picture of current selected receptors and exposures, these may be different from those that may be present and at-risk in the time period of several years to a few decades.

5. There is a failure to account for observational artifacts (such as time of nest initiation or failure) in the great horned owl study.

6. The MSU studies include confounding effects of habitat differences between the KRSS sites and the reference site (Fort Custer). In addition, the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge (box use is known to be significantly affected by familiarity of the birds with the placement of the boxes). These two issues significantly undermine the defensibility of conclusions drawn from these studies.

7. The MSU studies included averaged exposures over selected sampling sites within the study areas, rather than developing spatially explicit models of uptake and exposure. This is especially important for species whose foraging ranges are small relative to the scale of the study areas. MSU did not develop probabilistic exposure distributions representing

the relative variability in exposure over space. However, the Panel believes that such analyses may be possible with the MSU data.

8. The Panel believes that selection of the TRVs following a statistical approach that incorporates all available relevant toxicity data is a preferred approach to selecting the TRV based on a single study, in part because that approach addresses issues of cross-species extrapolations. This is important both because the species of concern may be more sensitive than some tested species, and because the trophic structure of the ecosystems may change over time, resulting in different species of concern. The Panel notes, however, that the single-study method is also a USEPA-approved methodology.

9. There is inadequate documentation and justification of the selected TRVs. The Panel strongly feels that because the selection of the TRV values is critical to determining the results of HQ-based risk assessments, there must be adequate documentation and justification of the data and the process used to derive the TRVs. If MSU is to base a TRV on a selected individual toxicity study instead of a statistical approach, then the rationale for selecting the particular study as the basis for the TRV and for rejecting other studies needs to be provided. This will avoid any *perception* that the assignment of the TRV used in Hazard Quotient calculations involved selective cherry-picking from available datasets. However, the particular TRV selected for shrews seems appropriate and representative of the available data. Thus, the point here is not that the TRV values used by MSU are necessarily incorrect, but, rather, that their derivations need further justification, and, since the selection of a TRV directly affects the results of the risk assessment, extra care needs to be placed on both the reality and the perception of an unbiased process.

10. The Panel noted several inconsistencies and incomplete explanations of methods in some of MSU's papers and Standard Operating Procedures.

11. There are conflicts between the statistical methods used by MSU and the methods prescribed in the Sampling and Analysis Plan and the Standard Operating Procedures provided to the Panel. For example, the SOPs outlined stratified random sampling, whereas subjective selection of sampling sites was used in the field studies.

12. The MSU studies overuse or inappropriately use results of tests of null hypotheses to pool data sets and reach conclusions. There is a general lack of quantification of sources of uncertainty in exposure and effects metrics. This is essential information for a decision-maker to use in a weight-of-evidence-based approach.

13. The MSU studies failed to address plausible future scenarios of the environmental conditions at the former impoundment sites; examples include removal of the remnant dam structures, as planned by the State, the potential effects of climate change, and an understanding of habitat succession at the sites of concern. Such changes could directly alter the trophic structure that determines exposure pathways and therefore risks.

14. The species studied (bluebirds, wrens, and owls) do not necessarily represent the most highly exposed or the most sensitive species present in the riparian corridor. Therefore, they are not adequate surrogates for addressing the question of risk to all avian species that may potentially use the site. The value of the great horned owl as a suitable receptor

species is limited by its large home range relative to the spatial scale of the impoundments (and the reference area). With only a few potential nesting territories present in each area, relatively short-term studies (i.e., several years) offer limited representativeness of the ecological, reproductive, and exposure-related variables affecting this species.

The following items summarize remaining uncertainties that should be resolved as the risk-management process on the KRSS moves forward. The data collected by MSU may help to reduce many of these uncertainties; however, none has been directly addressed in the published papers.

1. Uncertainty concerning the importance of the earthworm pathway — No worms were found in the diets of bluebirds and house wrens. Some earthworm data were presented from shrew studies, but these have not been fully examined in the context of a comprehensive conceptual model of risk pathways. Hence, in spite of the critical importance of the earthworm-to-robin pathway shown in the BERA, understanding the risks associated with the earthworm pathway is still unresolved especially for birds. Information is also needed on the locations of the earthworm samples relative to the soil samples used to derive BSAF values. These locations should also be compared to the range of soil values in the Trowbridge Impoundment.

2. Uncertainty concerning population-level effects — The nest productivity studies conducted by MSU address organism-level effects but do not support inferences concerning population-level effects of PCB exposures.

3. Uncertainty concerning assessment of passerine reproductive productivity — MSU's analysis of reproductive success would have been enhanced significantly by the use of standard methods (e.g., the Mayfield and related modern methods) and an integrated measure of fledging rate based on the number of all nests initiated. MSU's analysis partitioned reproductive success into various subcomponents and used only subsets of nests for some measures (e.g., the "predicted number of fledglings" used the smallest subset of nests, and "fledging success" ignored nests in which no eggs hatched). This approach potentially underestimates the cumulative effects of nest failure, embryonic mortality, and pre-fledgling mortality. For example, the Panel recalculated an overall measure of reproductive success by multiplying clutch sizes (# of eggs laid/initiated nest) by productivity (number of fledglings/egg laid), resulting in a 47% lower estimate of fledglings per nest initiated for bluebirds and 18% lower for house wrens at Trowbridge compared to the reference area. The full value of MSU's reproductive data is uncertain until a more thorough reanalysis is conducted along the lines noted above.

4. Uncertainty concerning extrapolation to other species — The species chosen for study by MSU were selected based on amenability to study and are not necessarily representative of typical species utilizing the site or of the most highly exposed or the most sensitive species. In particular, there is still substantial uncertainty concerning risks to ground-

feeding birds such as robins and woodcocks. The utility of MSU's PCB exposure data for robins is uncertain because although the birds were collected within the Trowbridge floodplain, observational data suggested that robins were foraging primarily outside of the floodplain. In the future, as soils develop and inundation decreases, earthworm populations and consequently feeding by robins are likely to increase within the floodplain. House wrens had elevated PCB exposure while feeding primarily on terrestrial invertebrates, but the lack of earthworms in their diet does not address the potential for earthworms as a key exposure pathway for birds such as robins and woodcock. The MSU data on earthworms at Trowbridge showed substantially higher PCB concentrations than other terrestrial invertebrates (Blankenship et al. 2005), further emphasizing the importance of assessing the earthworm exposure pathway.

5. Uncertainty concerning extrapolation to other sites — Types of habitats and plant communities found in the former impoundments appear similar, but the impoundments differ with respect to the relative size and spatial distribution of habitat types. Detailed comparisons of within-species differences in diets of bluebirds and house wrens foraging in Trowbridge vs. Fort Custer could be used to bound the possible range of diet variation among the other impoundments. The habitat characteristics of the Fort Custer site differs so much from the former impoundment sites that its utility as a reference site is quite limited.

6. Uncertainty concerning extrapolation to future conditions — Key habitat characteristics in all of the former impoundments can be expected to change over time from several plausible causes, including normal ecological succession, climate change, and a State-planned removal of remnant dam structures at Trowbridge, which will likely change the hydrological regime of the site and result in a significantly altered set of exposure pathways. It may be possible to use the data from the MSU diet studies, together with predictions of future habitat conditions, to predict future diets and exposure levels.

The following items summarize the Panel's **major** recommendations for using the results of the MSU studies in future risk assessment and risk management activities on the KRSS.

1. The Panel recommends that the **MSU studies conclusions** not be used to reach risk conclusions on their own. There is too much uncertainty underlying the data interpretation, lack of robustness in the study design, and insufficient documentation (and lack of agreement) of TRV derivations. However, the **MSU data** when combined with data from the BERA can be useful to inform the ecological risk assessment and risk management decisions associated with the KRSS.

2. The Panel recommends that, given the complexity of the datasets developed by USEPA and MSU, a multi-party technical working group consisting of scientists representing USEPA, MDEQ, and KRSG should be established to oversee the conceptual model development, cross-comparisons, uncertainty analyses, and dataset interaction activities

needed to integrate the MSU data with the dataset used for the BERA. Further, the Panel recommends that any future ecological risk assessment activities performed to support remedial actions at the Kalamazoo River site should be developed cooperatively, using a Data Quality Objectives (DQO) process based on USEPA guidance and other applicable documents. Following the DQO approach, the technical working group would develop a comprehensive conceptual model and identify key receptors of concern for the ecological risk assessments.

3. The Panel recommends that the parties involved collaboratively develop a comprehensive conceptual ecological model (CEM) that captures the exposure pathways and species or other ecosystem attributes of concern. The purpose of such a CEM is to capture the scientific understanding of the ecosystem and the stressors impinging on it. A well-designed CEM also identifies the important pathways contributing to the risk and the uncertainties associated with each pathway. The Panel believes that this is a critical step that needs to be accomplished before decisions can appropriately be made based on the best available science concerning the Kalamazoo restoration. The Panel recommends that in developing the CEM, the parties involved focus primarily on capturing the scientific understanding of how this ecosystem works and how it is structured, and not begin with a focus on the data available or studies that have already been done. The Panel also recommends that the CEM explicitly incorporate a coupling of the aquatic and terrestrial systems into an integrated CEM. The Panel also recommends that outside expertise be recruited into the CEM-development process.

4. The Panel recommends using a systematic approach incorporating all data from both MSU and BERA. Suggested activities include the following: (1) evaluation of the ability to pool the MSU and BERA data into a unified data set; (2) evaluation of the methods used to quantify the magnitude of PCB concentrations generated by each study (e.g., are the analytical chemistry results consistent among studies?); (3) calculation of uncertainty in the BERA results based on the extended data generated by MSU; (4) re-evaluation of the MSU statistical results using alternative statistical estimates of central tendency of measures of PCB concentrations; (5) comparison of BERA and MSU results using formal uncertainty analysis methods; (6) evaluation of the effect of temporal variability on the MSU and BERA findings; (7) comparison of the BERA findings to model-based findings that could be generated using the MSU data (see Appendix A); (8) evaluation of the findings in light of the conceptual model employed by each study; and (9) evaluation of the BERA findings in light of the quantifiable relationships between PCB source concentrations and PCB egg and body burden concentrations that may be obtainable with the MSU data (note: these analyses were not implemented by MSU, but the MSU data set suggests that such analyses may be possible).

5. The Panel further recommends that the risk assessment process include a set of scenario-consequence analyses, in which a series of plausible future conditions are incorporated into the CEM and the assessments are done on the resulting risks. A nested approach is recommended, in which the hydrological regime at present is the basis for examining a few scenarios of trophic structure and associated species of concern that exist now and at selected points-in-time as succession proceeds. Subsequently, an alternate hydrological regime should be considered, such as following removal of remnant dam structures, and associated stages in succession under the altered hydrological regime be assessed.

6. As an initial point-of-departure, it would be informative to perform a cross-comparison between MSU and BERA studies, by using the MSU data and assumptions as inputs to the BERA models and the BERA data and assumptions as inputs to the MSU models; an example of this is included in the Panel's Report, in which BERA and MSU exposures and TRV values are interchanged. It is to be expected that the results will differ between the two studies because of differences in scopes and study designs. Nevertheless, such comparisons could illustrate the magnitude of differences resulting from the two approaches and provide insights into the causes of the differences. Such a comparison would inform risk managers concerning validity and defensibility of each set of analyses. This exercise would seem to be essential before any reasonable understanding of the multiple-lines-of-evidence approach could be reached from the two disparate datasets and results. Following this initial exercise, a more systematic approach to using all of the available data from both MSU and USEPA is recommended; the Panel outlines specific steps in such analyses in the Report.

7. The Panel recommends that rather than focus on estimating a single risk number, the ecological risk assessments would be strengthened by presentation of a distribution of risk levels tied to the uncertainties in the underlying data and/or model structure (e.g., relative importance of different dietary pathways). Consequently, the Panel recommends that the exposure models and data in the MSU study and the BERA be subject to a formal uncertainty analysis. The scope of the formal uncertainty analysis could be informed by results of a bounding analysis. Included in the uncertainty analysis should be an extensive sensitivity analysis of the models to explore the plausible range of risks in the system.

8. The Panel recommends that consideration also be given to applying an approach to selecting TRVs that uses the full set of available, high-quality toxicity studies rather than relying on a single study to derive a TRV. Some possible approaches are suggested in the Panel's Report. Regardless of what method is used, the Panel recommends that the derivation of the TRVs is explicitly described and documented, and the specific TRVs selected be fully justified to enhance the confidence that the TRVs that are selected are appropriate and protective. A more comprehensive approach would be to use a range of plausible TRV values for each receptor of concern, enhancing the utility of the results to the multiple-weight of-evidence approach for risk management.

Additional Panel recommendations can be found in section **5.0 Recommendations** of the Report.

2.0 Introduction

2.1 Background

Following the release of the initial Baseline Ecological Risk Assessment (BERA), the Kalamazoo River Study Group (KRSG) provided a series of independent grants to Michigan State University (MSU) for additional ecological studies. The Final Revised Baseline Ecological Risk Assessment (CDM 2003) was finalized before these studies were completed. In February 2007 KRSG voluntarily entered into an Administrative Settlement Agreement and Order on Consent (AOC) for the Site.

The AOC describes a series of supplemental RI/FS activities, potentially including completion of Area-Specific Ecological Risk Assessments. The KRSG requested that the MSU studies be considered as additional lines of evidence for evaluating ecological risks and for subsequent risk management decisions conducted as part of the AOC.

The Statement of Work (SOW) attached to the AOC called for the MSU studies pertaining to floodplain soils to be subjected to a peer-review process prior to incorporation. Floodplain refers to the areas of formerly impounded sediments (i.e., the extent of inundation prior to the lower of water levels in the impoundments). Consequently, the peer review focused on these exposed sediments. Dr. John Giesy of Michigan State University (MSU) led the studies on the exposure and effects of PCBs in floodplain soils in the former Trowbridge Impoundment and a reference site (Fort Custer) under independent grants from KRSG. The studies included:

- Productivity assessment of two passerine species and great horned owl;
- Measures of dietary composition for birds;
- Measures of prey tissue PCB concentrations for birds;
- MSU's ecological risk assessment.

The Peer Review Panel (Panel) used the *Final (Revised) Baseline Ecological Risk Assessment Allied Paper, Inc. /Portage Creek/Kalamazoo River Superfund Site* (Site-Wide Baseline ERA; CDM 2003) to provide context and inform the review, but the Panel was not charged with reviewing this document.

However, the Peer Review Panel was charged with independently reviewing the MSU studies and responding to six (6) charge questions, each with supplemental issues noted, and to support answers with citations or other background information as appropriate.

2.2 Peer Review Process

The peer review process was initiated collaboratively by the Kalamazoo River Study Group representing the PRPs and USEPA by agreeing to engage Dr. Ken Dickson to be the Peer Review Manager. KRSG with the assistance of ARCADIS Inc. and the USEPA prepared independent lists of potential scientists to participate on the Peer Review Panel. Dr. Dickson was requested to constitute the panel but not be constrained by the lists provided. In March 2008 Dr. Dickson recommended to KRSG and USEPA a list of seven (7) scientists to be on the panel. KRSG and USEPA accepted all of the recommended scientists. The Peer Review Panel was constituted in April 2008. The peer review process began on May 13 at a Charge Meeting held in Augusta, Michigan. Participants at the Charge Meeting included representatives from KRSG, ARCADIS Inc., USEPA, MDEQ, MDNR, NOAA, CDM Inc., and MSU. At the meeting Dr. Ken Dickson, the Peer Review Panel, and other participants were introduced to the peer review process to be followed and the project schedule. Dr. Ken Jenkins (ARCADIS Inc.) and Dr. James Chapman (USEPA) provided general background information, introduced the charge to the Panel, and answered questions. The Panel was then briefed by Camp Dresser and McKee (CDM) representatives on the Final Baseline Ecological Risk Assessment (BERA- CDM 2003). Dr. John Giesy, principal investigator of the MSU studies, presented a detailed summary of the MSU studies, results, and conclusions, and responded to questions from the Peer Review Panel and other meeting participants. Representatives from USEPA, MDEQ, and MDNR offered comments/questions about the MSU studies and were requested to provide written comments to the Panel. At the conclusion of the meeting, the Peer Review Panel accepted the charge. The following day, May 14, 2008, the Peer Review Panel and representatives from KRSG, USEPA, and MDEQ visited the study sites used for the MSU studies at the Trowbridge Impoundment and Fort Custer. The group also visited other Kalamazoo River Super Fund Sites (KRSS) along the river.

The Peer Review Panel was provided the following information/documents to inform the review:

- Final (Revised) Baseline Ecological Risk Assessment Allied Paper Inc./Portage Creek/ Kalamazoo River Superfund Site, Michigan Department of Environmental Quality Remediation and Redevelopment Division, April 2003

- Ecological Consequences of PCBs in the Exposed Sediments of Formerly Impounded Areas of the Kalamazoo River – Overview of Studies Conducted by Michigan State University Prepared by Dr. John Giesy and Dr. Matthew Zwiernik – Prepared on behalf of the Kalamazoo River Study Group, May 2008

- The following nine published papers resulting from the Michigan State University studies of the fate and effects of PCBs in the Kalamazoo River:

    - Blankenship, A.L., M.J. Zwiernik, K.K. Coady, D.P. Kay, J.L. Newsted, K. Strause, C. Park, P.W. Bradley, A.M. Neigh, S.D. Millsap, P.D. Jones, and J.P.

Giesy. 2005. Differential accumulation of polychlorinated biphenyl congeners in the terrestrial food web of the Kalamazoo River Superfund site, Michigan. Environmental Science and Technology 39:5954-5963.

- Giesy, J. and M. Zwiernik. 2008. Ecological consequences of PCBs in the exposed sediments of formerly impounded areas of the Kalamazoo River. Presentation to the Kalamazoo River Ecological Risk Studies Peer Review Panel, 13 May 2008.

- Neigh, A.M., M.J. Zwiernik, A.L. Blankenship, P.W. Bradley, D.P. Kay, M.A. MacCarroll, C.S. Park, P.D. Jones, S.D. Millsap, J.W. Newsted, and J.P. Giesy. 2006a. Exposure and multiple lines of evidence assessment of risk for PCBs found in the diets of passerine birds at the Kalamazoo River Superfund site, Michigan. Human and Ecological Risk Assessment 12:924-946.

- Neigh, A.M., M.J. Zwiernik, C.A. Joldersma, A.L. Blankenship, K.D. Strause, S.D. Millsap, J.L. Newsted, and J.P. Giesy. 2007. Reproductive success of passerines exposed to polychlorinated biphenyls through the terrestrial food web of the Kalamazoo River. Ecotoxicology and Environmental Safety 66:107-118.

- Neigh, A.M., M.J. Zwiernik, P.W. Bradley, D.P. Kay, P.D. Jones, R.R. Holem, A.L. Blankenship, K.D. Strause, J.L. Newsted and J.P. Giesy. 2006a. Accumulation of Polychlorinated Biphenyls (PCBs) from Floodplain Soils by Passerine Birds. Environmental Toxicology and Chemistry, Volume 25, pp. 1503-1511.

- Strause, K.D., M.J. Zwiernick, S.H. Im, P.W. Bradley, P.P. Moseley, D.P. Kay, C.S. Park, P.D. Jones, A.L. Blankenship, J.L. Newsted, and J.P. Giesy. 2007a. Risk assessment of great horned owls (*Bubo virginianus*) exposed to polychlorinated biphenyls and DDT along the Kalamazoo River, Michigan, USA. Environmental Toxicology and Chemistry 26:1386-1398.

- Strause, K.D., M.J. Zwiernick, S.H. Im, J.L. Newsted, D.P. Kay, P.W. Bradley, A.L. Blankenship, L.L. Williams, and J.P. Giesy. 2007b. Plasma to egg conversion factor for evaluating polychlorinated biphenyl and DDT exposure in great horned owls and bald eagles. Environmental Toxicology and Chemistry 26:1399-1409.

- Strause, K.D., M.J. Zwiernik , J.L. Newsted, A.M. Neigh, S.D. Millsap, C.S. Park, P.P. Moseley, D.P. Kay, P.W. Bradley, P.D. Jones, A. L. Blankenship, J.G. Sikarskie, and J.P. Giesy. 2008. Risk assessment methodologies for exposure of great horned owls (*Bubo virginianus*) to PCBs on the Kalamazoo River, Michigan. Integrated Environmental Assessment and Management 4:24-40.

- Zwiernik , M.J., K.D. Strause, D.P. Kay, C.S. Park, A.L. Blankenship, and J.P. Giesy. 2007. Site-specific assessments of environmental risk and natural resource damage based on great horned owls. Human and Ecological Risk Assessment 13:966-985.

- Michigan State University Sampling and Analysis Plan (SAP) and Standard Operating Procedures (SOPs) dated January 7, 2000
- Michigan State University Quality Assurance Project Plan for the Kalamazoo River Area of Concern Baseline Ecological Risk Assessment dated January 7, 2000
- MSU's Kalamazoo Data Base
- NOAA Kalamazoo Data Base
- ARCADIS. 2008. Characteristics of the Formerly Impounded Areas. April 2008.

In addition to the above sources of information the Peer Review Panel requested and received the following:

- MSU's shrew data and a report discussing the results (MSU, 2001);
- MSU's explanation of the process followed to choose the TRVs used in their analyses;
- Information from Michigan Department of Natural Resources and Michigan Department of Environmental Quality on potential future land use management of the formerly inundated floodplains at the sites.
- MSU's spatial data on soil, earthworm and shrew PCB levels in sampling grids for Trowbridge and Fort Custer

The Panel reviewed the results of the MSU studies in June, July, August and September and prepared a Draft Final Report in September, 2008. The Draft Final Report was reviewed by USEPA, MDNR, and KRSG. Written comments and questions about the Draft Final Report were provided to the Panel in late September and a one day meeting was held September 25 in Detroit where the comments and questions received were discussed by the Peer Review Panel with representatives from USEPA, MDNR, and KRSG. Following the September 25, 2008 meeting, the Peer Review Panel developed this Final Report giving consideration to all input received.

2.3 Summary of KRSG and USEPA's Charge to the Panel

2.3.1 General Guidance to Panel Regarding the Charge

The charge to the Peer Review Panel was to review the MSU studies with respect to their suitability as additional lines of evidence for evaluating potential risks to terrestrial receptors exposed to PCBs in floodplain soils in the formerly impounded areas of the Kalamazoo River. A summary of the MSU studies and supporting information was provided to assist the Panel understanding the material to be reviewed. The Panel was also asked to review the Baseline ERA (CDM 2003) for important supporting information and lines of evidence for future risk management decisions. The Baseline ERA (CDM 2003) provided context for the MSU studies, which were designed to provide additional lines of evidence for consideration in the final risk management decisions. However, the Baseline ERA (CDM 2003) was not peer reviewed by the Panel.

The primary objective of the peer review process was for the Panel to provide an independent technical opinion regarding the extent to which the information in the MSU studies could be incorporated as independent lines of evidence, along with those presented in the Baseline ERA (CDM 2003), in a weight-of-evidence evaluation of ecological risks to terrestrial receptor species in formerly impounded areas and for subsequent risk management decisions. In reviewing the materials associated with the MSU studies, the Panel was requested to weigh the following general questions when addressing the specific questions presented in the charge:

1. Are the methods employed in the MSU studies appropriate and consistent with the current state of the science and relevant guidance?
2. Have uncertainties associated with the MSU studies been clearly identified and discussed?
3. Do the data and analyses presented in the MSU studies constitute reasonable and appropriate lines of evidence to consider in the evaluation of risks to terrestrial receptors in future risk management decisions?
4. Do the MSU studies represent reasonable and appropriate lines of evidence for consideration in risk management decisions regarding the formerly impounded areas?

2.3.2 Specific Questions to be addressed by the Panel

2.3.2.1 Exposure Assessments

This section addresses specific issues regarding the evaluation and interpretation of levels of exposure to PCBs for receptors that use the floodplains of the formerly impounded areas. A summary of the types of data and strategies employed by MSU for the evaluation of exposure for the various receptor species is presented in Table 2.1 (included below). The Panel was asked to address the following question regarding exposure and the supplemental issues:

Question 1. What are the relative strengths, limitations, and uncertainties associated with the methods employed by MSU to estimate the exposure of each receptor species to PCBs?

Supplemental Issues to Consider:

1a. Relative strength of various measures of exposure evaluated for each receptor when available individually and in combination. Examples of the types of data MSU considered include: a) literature-,based information on preferred prey; b) Site-specific data on receptor-specific prey items; c) site-specific bioaccumulation factor-based estimates of PCBs in prey; d) direct measures of PCBs in prey; and e) direct measures of PCBs in tissues/eggs of receptors.

1b. Effects of differing dietary preferences on extrapolating from the results of the MSU studies to other species. As an example, how may species-specific dietary preferences of the wrens or bluebirds evaluated in the MSU studies affect extrapolation of risk from these species to robins?

1c. The potential effects of future conditions, such as possible changes in habitat over time due to natural succession or anthropogenic changes to enhance recreational use. Some examples include lowered water table and reduced soil moisture content related to dam removal, transition to meadows including short grass habitat or succession to mature hardwood forest.

Table 2.1 Data Types Available for Refining PCB Exposure Estimates

Available Data	**Proposed Use**
Bird tissue data presented in Blankenship et al. (2005); Neigh et al. (2006b); Strause et al. (2007a, b, 2008); Zwiernik et al. (2007); and the summary of the MSU studies.	Develop estimate of avian body burden for use in dose model for upper trophic level species.
Shrew and other small mammal tissue data presented in CDM (2003), Blankenship et al. (2005), and the summary of the MSU studies.	Develop estimate of small mammal concentration for use in dose model for upper trophic-level species.
Invertebrate tissue data presented in CDM (2003), Blankenship et al. (2005), and the summary of the MSU studies.	Develop estimate of invertebrate concentration for use in dose model for insectivores.
Egg concentrations from multiple avian species presented in CDM (2003), Neigh et al. (2006b, 2007); Strause et al. (2007a); Zwiernik et al. (2007); and the summary of MSU studies.	Compare to egg-based TRV.
Great horned owl pellet analysis and passerine nestling dietary composition analysis conducted as part of the MSU studies (Neigh et al. 2006a; Strause et al. 2008; Zwiernik et al. 2007).	Refine estimate of dietary composition for purpose of dose modeling.

2.3.2.2 Effects Assessment

This section addresses specific issues regarding the strategies employed in the MSU studies to evaluate potential effects of PCB exposure on receptors utilizing the floodplains of the formerly impounded areas. The Panel was charged to address the following questions regarding effects and the supplemental issues:

Question 2: What are the relative strengths, limitations, and uncertainties associated with the productivity assessments conducted by MSU on passerines and great horned owls (Neigh et al. 2006a, 2007; Strause et al. 2007a, 2008)?

Supplemental Issues to Consider:

2a. Strengths and limitations of directly measuring productivity in the field compared to extrapolating from controlled laboratory studies.

2b. Extrapolation of results from field productivity studies to other species such as the American robin, which was the receptor species considered in the Baseline ERA (CDM 2003).

2c. Evaluation of potential causal factors (e.g., PCB concentrations, habitat differences, etc.) associated with any difference in measures of productivity in passerines relative to the reference site.

Question 3: What are the relative strengths, limitations, and uncertainties associated with the hazard quotient calculations performed by MSU to evaluate potential risk to passerines, great horned owls, and shrews (Neigh et al. 2007; Strause et al. 2007a, 2008)?

Supplemental Issues to Consider:

3a. Choice of toxicity reference value (TRV), including relevance to receptor species and quality of study (e.g., duration, inclusion of sensitive life stages, exposure range, endpoints measured).

3b. Uncertainty resulting from extrapolating from laboratory study to field.

3c. Uncertainties in extrapolating from one species to another.

2.3.2.3 Applicability of the Investigations

This section of the charge addresses the overall quality of the data and the analyses presented in the MSU studies and their applicability for the evaluation of ecological risk and supporting risk management decisions for the floodplains of the formerly impounded areas. With this in mind the Panel was asked to address the following questions:

Question 4: What are the relative strengths, limitations, and associated uncertainties that should be considered when evaluating the results of these studies as potential lines of evidence in future risk management decisions?

Supplemental Issues to Consider:

4a. Study designs including (but not limited to) sample size, replication, temporal duration, and aggregation of data.

4b. Data interpretation, including the choice and application of statistical methods.

4c. Approach for addressing natural variability.

4d. Identification and characterization of uncertainties.

4e. Adequacy of the data to support inferences on population-level effects.

Question 5: What are the relative strengths, limitations, and associated uncertainties that should be considered when extrapolating from the results of MSU studies conducted in the former Trowbridge Impoundment to the other formerly impounded areas of the Kalamazoo River?

Supplemental Issues to Consider:

5a. Numeric and spatial distributions of PCBs in floodplains of former impoundments.

5b. Habitat characteristics in floodplains of formerly impounded areas.

5c. Likely utilization of floodplains in formerly impounded area by the receptor species evaluated in MSU studies

2.3.2.4 Risk Management

This section of the charge addresses the potential usefulness of the MSU studies in supporting risk management decisions for the floodplains in the formerly impounded areas. It is possible that the results of the MSU studies would be incorporated as independent lines of evidence, along with data from the Baseline ERA (CDM 2003), in an Area-specific ecological risk assessment process. With this in mind please address the following question.

Question 6: Please comment on the applicability of the information presented in the MSU studies for informing risk management decisions.

3.0 Panel's Responses to Charge Questions

3.1 General Charge Questions

3.1.1 Are the methods employed in the MSU studies appropriate and consistent with the current state of the science and relevant guidance?

Panel's Response*:* The MSU studies encompass significant field studies measuring exposure and effects as well as risk assessment models for estimating both, reflecting the current state-of-the-science. This combined strategy has several advantages, including the potential to validate modeled predictions with real-world data for the several species that were actually studied, and the ability to use PCB concentrations in soil and food items to model both exposure and effects for species that were not studied in the field. The general design and execution of the field studies were comparable in quality to other field assessments of PCBs in riparian habitats in large river systems. However, the methods used by MSU for assessing passerine productivity and for selecting TRVs were not the current state-of-the-science. Another limitation was the use of non-rigorous statistical design and analysis of results.

3.1.2 Have uncertainties associated with the MSU studies been clearly identified and discussed?

Panel's Response: The MSU Summary Report and publications resulting from the study do not adequately identify or explain the uncertainties associated with exposure, effects, ecological risks, and risk management conclusions. A **strength** of the MSU field studies was that they provided site-specific data, thereby reducing uncertainties inherent in the approach of using primarily literature-based values to estimate exposures and effects. However, the small sample sizes, issues related to the adequacy of reference sites and statistical analyses, and absence of information on some important pathways of exposures **limit** the ability of the field-based approach to reduce uncertainties.

An approach to better explain (and quantify) uncertainties would be to examine the exposure model(s) in the MSU study and BERA using formal uncertainty analysis to explore the plausible range of risks in the system. For example, one of the variables in such a model is the particular diet of an endpoint species. A probability distribution (consistent with field measurements) of dietary sources could be generated allowing quantification of the uncertainty in exposure as a function of diet. Similarly, use of different specific bioaccumulation factors within the range of plausible values for each could be explored in a set of Monte Carlo simulations using hazard quotients. In general, rather than focus on deriving a single hazard quotient, the risk assessments would be strengthened by presentation of a distribution of risk levels tied to the uncertainties in the underlying data or model structure (e.g., relative importance of different dietary pathways).

3.1.3 Do the data and analyses presented in the MSU studies constitute reasonable and appropriate lines of evidence to consider in the evaluation of site-specific risks to terrestrial receptors in area-specific risk assessments?

Panel's Response: It is the Panel's opinion that data developed by the MSU studies have the potential to inform the site-specific risks (that is, information about the Trowbridge impoundment) to terrestrial receptors, but the study design and analyses of data were inadequate, and in some cases inappropriate, to inform area-specific risk assessments (e.g., extrapolation to the other impoundments or the entire contaminated stretch of the river). The MSU data analyses do not take full advantage of the information contained in the data sets, and therefore do not fully explore the lines of evidence that are inherent in the collected information. Thus, while the MSU studies do present information that can reasonably contribute to a multiple-lines-of-evidence assessment, these data, results, and conclusions need to be considered with caution and appropriate recognition of their uncertainties and limitations.

3.1.4 Do the MSU studies represent reasonable and appropriate lines of evidence for consideration in risk management decisions regarding the former impounded areas?

Panel's Response: The results of the MSU studies should be included in a multiple-line-of-evidence approach for risk assessment and risk management decision-making, with caution and appropriate recognition of their uncertainties and limitations. The MSU data and the BERA data should be used to develop an integrated multiple-lines-of-evidence-based ecological risk assessment (i.e., possibly using data from both studies in a single data analysis approach) to inform risk management decisions for the formerly impounded areas. However, the uncertainties associated with the multiple lines of evidence from the BERA and MSU studies should be identified and formally quantified so that they can be more effectively considered and weighed in the risk-management process.

3.2 Specific Charge Questions Addressed by the Panel

3.2.1 Question 1: What are the relative strengths, limitations, and uncertainties associated with the methods employed by MSU to estimate the site-specific exposure of each receptor species to PCBs?

Panels' Response: The **strengths** of the MSU studies lie in their direct measurement of PCB concentrations in prey items, soil, receptor species, and actual diets of receptors of choice for the field studies. Such measurements can significantly reduce the uncertainty of exposure estimates for the receptors of concern. The **limitations** of the MSU studies include small sample sizes (see Appendix A for further remarks on small sample sizes) for some species/trophic levels, lack of spatially explicit data (e.g., BAF functions cannot be determined, only BAF constants), and measurement of PCB concentrations as wet weight (which adds significant variability as opposed to reporting out as dry weight). These limitations result in continued uncertainties in dietary estimates of food chain-modeled species. Additionally, a number of other **limitations** lead to **uncertainties** in MSU's risk conclusions, as well as in extrapolations across habitat types, impoundments, and time periods. MSU's papers are limited in their reporting of locations and habitat types associated with samples. Small sample sizes for some variables reduce the potential for reanalysis of the data on a spatial basis. In other cases, the compositing of data across sampling areas and habitat types has reduced the level of resolution of the exposure analysis. The study design was limited to assessing current risks to a specific set of receptors, and failed to adequately address the robin pathway that was identified as critical in the BERA. The analysis of PCB weathering is flawed in the calculation of relative potency factors for irrelevant exposure pathways (e.g., MSU's data show that shrews and other small mammals constitute a small fraction of the owls' diet, hence preliminary remediation goals (RPGs) meant to represent weathering of congener mixtures along this pathway are meaningless). Other limitations and uncertainties are identified in the Panel's responses to specific questions that follow.

Supplemental Issues to Consider in Question 1:

> ***1a. Relative strength of various measures of exposure evaluated for each receptor when available individually and in combination. Examples of the types of data MSU considered include: a) literature-based information on preferred prey; b) Site-specific data on receptor-specific prey items; c) site-specific bioaccumulation factor-based estimates of PCBs in prey; d) direct measures of PCBs in prey; and e) direct measures of PCBs in tissues/eggs of receptors.***

Panel's Response 1a and 1b: Site-specific diet data generally are preferable to literature-derived diets. The results from the MSU studies could be used to calculate species-specific diets used in the area-specific ecological risk assessments. However, differences in diets between impoundments, as influenced by habitat characteristics, site history, and other factors, must be taken into account to the extent possible. Many birds, in particular here robins, spend a substantial amount of their foraging time off-site. Exposure estimates for these species should be adjusted to account for off-site foraging. Similarly, some species such as shrews have very small home ranges, so exposure estimates should not be averaged across an entire impoundment. MSU's data also show that a substantial fraction of the diets of bluebirds nesting in the former impoundments consist of aquatic insects, and the great horned owl diets also have a significant amount of items from the aquatic food web (Strausse et al. 2008). Therefore, the potential contribution of PCB exposure from aquatic insects should be considered in area-specific risk assessments of these (or similar) species. Supporting references are: Neigh et al. (2006) and Strausse et al. (2008).

Panel's Response to Question 1c – Strengths of Measures of Exposure: Site-specific data on bioaccumulation are nearly always preferable to literature-based bioaccumulation factors;

however, site-specific BAFs are not necessarily transferable even between nearby sites. Bioavailability of PCBs in soils may differ, and the utilization of the site by target receptors may also vary between sites. When extrapolating to other impoundments, it is necessary to identify the uncertainties related to possible variations in PCB bioavailability and receptor dietary preferences. Because the biota-soil accumulation factors (BSAFs) are carbon-normalized, extrapolation among sites is more reliable if there are data on soil carbon for all areas. Similarly, differences in PCB congeners at the various sites should be taken into account when doing such extrapolations. Supporting Reference: Blankenship et al. (2005) contains site-specific biological accumulation factors (BAFs) and biological magnification factors (BMFs) calculated from the Trowbridge impoundment data for birds.

The Panel is concerned that the method used to determine the BAF factors may result in inaccurate "safe" values in the soil when calculating PRGs for clean up. Note that PCB concentrations in soil are expressed on a dry-weight basis, whereas tissue concentrations are all on a wet-weight basis (or lipid-normalized, but still wet-weight). Since the percent moisture in any of the biota or tissue samples is not known (or at least is not reported), the actual ***mass*** of PCBs that has moved up the food chain is uncertain; the relative concentrations are confounded by the differences in percent moisture among individuals and between species. In addition, the method for preparing biota for PCB analysis may change their hydration, thus altering their measured wet weight. For example, earthworms were depurated prior to chemical analysis by placing them on filter paper for several hours. During this time, they likely desiccate to some extent; thus, their measured "wet weight" is not a true field weight and more variability is added to the analysis (plus individual earthworms will desiccate to different degrees). Although this method of calculating BAFs on a wet-weight basis has become the standard approach for terrestrial systems (but not for aquatic food-chain analyses), it continues to be a major flaw in the way terrestrial food-chain risks and safe soil values are estimated.

Panel's Response 1c – Use of MSU Site-Specific Tissue Data in BERA Dietary Exposure Models: The **strength** of this approach is to provide site-specific BSAFs and BMFs and measured concentrations in biota. This can incorporate soil-specific effects (e.g., soil carbon), congener-specific differences in accumulation rates, and species-specific information related to the site (particularly for raptors, where literature-based data are very sparse). A **limitation** is that the robin diet was not specifically modeled, so it may be that all appropriate biota were not sampled. However there are some data for earthworms, so some estimate of site-specific risk could be developed using the dietary model in the BERA. Also, the recently provided shrew data could be used for site-specific exposure estimations. Background information in the BERA stated that robins can consume up to 90% of their diet as invertebrates during the breeding season, and only up to 20% invertebrates in the remainder of the year. The BERA assumed 51% of the robin's diet is from soil invertebrates. The USEPA Wildlife Exposure Factors Handbook (USEPA 1993) identifies 59 – 71% invertebrates in the robins spring/summer diet. No explanation is provided for why the BERA selected 51% as the amount of soil invertebrates ingested. Robins actually eat a wide variety of soil invertebrates, but as only the earthworm was measured for PCB concentrations, it was considered a "worse case" for PCB uptake and could be used in a site-adjusted estimate of dietary exposure to robins.

Panel's Response to 1b and 1d – Use the "Bolus" Data from the Avian Nesting Study to Further Verify Dietary Exposure Estimates: The **strength** of this approach is that the food bolus represents precisely what the nestlings are eating. By comparing the concentrations in this bolus to the estimated concentrations from the dietary exposure model, the model can be further refined to accurately reflect the diets and exposures of the studied species. This may provide some additional realism for extrapolating to the non-measured species, such as the robin. The **limitation** of this approach is that of the sampled birds, only the house wren is truly feeding on only terrestrial foods, whereas the eastern bluebird, the tree swallow, and the great horned owl access some (or most) of their diets from the aquatic food chain. Thus, relating diet to soil contamination alone is difficult, resulting in substantial uncertainty.

Panel's Response to 1b and 1d – Great Horned Owl (GHO): The prey item sampling (pellets and prey remains) was a critical part of the ecological studies that provided dietary composition data for bottom-up modeling of PCB exposure. Such studies must pay particular attention to prey identification and quantification in order to avoid biases that over- or under-estimate the frequency of particular food items. There are inconsistencies and incomplete explanations in MSU's descriptions of GHO prey item sampling and analysis. The SOP (273) and two published studies (Strause et al. 2007b, Zwiernik et al. 2007) vary somewhat in the citations for the prey analysis methods. Strause et al. (2007b) appears to have the best description of MSU's actual methods, although more details would be helpful. The methods actually employed differ significantly from those in the SOP, especially with respect to the schedule for collection (4X/month in the SOP vs. 2X per breeding season in the actual study) and level of data aggregation (individual pellets in SOP vs. composite samples in the study). In some cases the cited papers do not appear to provide strong support for the particular point that is being made by MSU (e.g., Hayward et al. 1993 in Zwiernik et al. 2007). While overall the estimated dietary composition methods appear to have provided satisfactory data, the inconsistent and incomplete documentation of methods potentially limits quantitative comparisons to other studies (which may have used other methods) and restricts the ability of other researches to replicate these protocols. The **strength**s of this study are the direct measurements of PCBs in prey items that rarely are analyzed, and a reasonable comparison between KRSS and the reference site. Further **strengths** are the presentation of data on both a mass-basis and a concentration-basis, plus inclusion of both means and 95% UCLs (upper confidence limits) of the means (although the means are geometric means, rather than the more appropriate means based on a lognormal distribution).

Panel's Response – Approaches for Calculating Soil PRGs for Receptors Exposed to Both Sediment-Derived and Soil-Derived PCBs: USEPA requested that the Panel provide guidance on approaches that have been used in other remediation activities to calculate soil PRGs when receptors are exposed to both sediment-derived and soil-derived PCBs.

MSU's studies demonstrated that a number of receptors feed on a mixture of aquatic and terrestrial prey. If a remediation goal for sediment has already been established, then soil PRGs for receptors exposed to both sediment-derived and soil-derived PCBs can be established by: (1) quantifying the relative contributions of aquatic and terrestrial food sources for that receptor; (2) estimating PCB concentrations in aquatic food sources, assuming that the sediment PRG is met; (3) establishing the acceptable daily dose (ADD) to the receptor(s) of concern; and (4) calculating the soil concentration corresponding to the remaining portion of the ADD.

If remediation goals for sediment have not yet been determined, then a matrix can be developed that provides pairs of potential sediment and soil remediation goals that would protect the receptors of concern. This would allow risk managers to explore alternative scenarios for sediment and soil remediation that would be equally protective of the receptors of concern.

Finally, PRGs for soil could be established using receptors with exclusively terrestrial-based diet.

Panel's Response to MSU Data Types Available for Refining PCB Exposure Estimates (Table 2.1 in Charge):

1. *Use of bird tissue data to develop estimates of avian body burden for use in dose model for upper trophic level species*

Panel's Response: Using the bird tissue data from the MSU studies would be preferable to using literature-derived bioaccumulation factors, but only if a sufficient number of samples exist for the particular species.

2. *Use of shrew and other small mammal tissue data to develop estimate of small mammal concentration for use in dose model for upper trophic level species*

Panel's Response: Use of shrew and other small mammal tissue could potentially provide a more realistic exposure estimate for terrestrial predators, if the sample sizes (number of animals and range of prey species) are sufficient to provide meaningful estimates of prey tissue concentrations. This is comparable to the work done by Strausse et al. (2008) for estimating great horned owl diets.

Panel's Response – MSU Shrew Studies: The MSU shrew study provides the following data:

- PCB tissue levels in shrews from the former Trowbridge impoundment; collections are from four grids and yielded 17 animals for tissue analysis.

- Trapping data from Trowbridge (4 grids) and the reference location at Fort Custer (2 grids); these data provided abundance caught relative to trap-nights (capture per unit effort), although trapping was terminated when a predetermined number of individuals were caught.

The Panel believes that these data could be used to evaluate exposures and risks to higher trophic-level predators, including risks to shrews, that is, the shrews should be considered both a pathway and receptor species of concern. Each of these is described below.

Panel's Response – Evaluating Exposures to Higher Trophic Levels: Shrews comprise part of the diet of various raptors and mammals. Therefore, the PCB body-burden data for shrews can be used to evaluate exposures to these trophic levels.

Panel's Response – Limitations/Qualifications of the MSU shrew studies to estimate exposures to higher trophic levels include:

- With respect to future use of the body-burden data, it will be important to examine the spatial relationships between the shrew tissue data and the relevant soil data at the scale of sampling grids. BAF values are not necessarily linear with soil concentrations. Therefore, the change in BAF(s) with soil concentrations can be important and should be examined prior to applying BAFs from one location to another.

- The Panel could not initially determine from MSU papers and reports which shrews come from which grids and how this relates to the associated soil PCB concentrations. An examination of these geographic and concentration relationships will be important to determine the feasibility of using these data to make extrapolations to other parts of Trowbridge (where shrews were not collected) and to the other former impoundments. (Note: The Panel received information November 26, 2008 on which shrew, earthworm and soil PCB samples came from which grid too late for a spatial analysis to be completed, vetted and included in this Report. However, the Panel will provide KRSG and EPA with a letter report on the results of this analysis).

Panel's Response – Evaluating Exposures to Shrews as Receptors: Measurements of PCBs in soils, earthworms, and shrews could be used to estimate PCB exposure and risks to shrews. Strengths of the MSU studies include:

- Direct measures of PCBs are available for a field-collected food item –earthworms – that likely reflect higher concentrations of PCBs than other potential invertebrate food items for shrews.

- If there is sufficient variability in soil concentrations among the soil-worm collection sites to develop a BAF function beyond a simple constant and associated uncertainties are considered, then the field-derived data could be applied to other areas. Otherwise, using the highest BAF could provide a conservative estimate.

- The combination of dietary exposure estimates and shrew body burdens provides a basis for a weight-of-evidence assessment of risk to shrews. However, the best use of the shrew data is as input to the food-chain model for higher-order predators.

Panel's Response – Limitations/Qualifications of MSU shrew studies include:

- The Panel could not easily discern the locations from which the earthworms came and the specific calculations used to associate earthworm tissue levels to soil concentrations. In order for the earthworm data to be of value, there needs to be a clear association of tissues with co-located soils. An examination of these geographic and concentration relationships is critically important to making use of these data for other parts of Trowbridge and for other former impoundments. Use of these data to make site extrapolations will need to be based on professional judgment because of sampling design, small sample size and pseudo-replication issues (See Appendix A).

- The sample size for earthworms is small, especially in light of heterogeneity in soil PCBs, and, therefore, care must be taken on how to extrapolate these data to other regions within Trowbridge and to other former impoundments. In particular, because BAFs may vary with soil concentration, this relationship should be examined as part of a data-usability assessment. If the range of soil concentrations used to derive earthworm BAFs for Trowbridge overlaps (is representative) of the range in other former impoundments, there is higher confidence in using the Trowbridge data. If the data sets are very different, there is less confidence. However, if other former impoundments have lower concentrations, the data for Trowbridge could still be used as a bounding analysis. Given the low numbers of earthworm samples, the Panel concurs with MSU to use the higher-derived BAF for extrapolations.

- The Panel recommends that these data be used as part of a weight-of-evidence approach and that the approach not rely on the HQ values that have been calculated. Instead, it is recommended that these types of calculations be repeated after the data have been processed as described above.

- The MSU shrew assessment relies on averages and upper bounds on averages for individual impoundments. The impoundments are large relative to foraging areas of shrews. Therefore, a spatially explicit approach should be used that considers the exposures distributions to individuals shrews across the impoundment(s). Exposure and risks can then be presented in terms of sub-areas or as fractions of the local population. The Spatially-Explicit Exposure Model (SEEM) offers one possible way for doing this.

3. *Use of invertebrate tissue data to develop estimate of invertebrate concentration for use in dose model for insectivores (birds? shrew?)*

Panel's Response: The available invertebrate tissue data could, in principle, provide more realistic exposure models for insectivorous birds and mammals, if the sample sizes are sufficient to provide meaningful estimates of prey tissue concentrations. It would be most appropriate to use invertebrate data in a spatially explicit analysis of the impoundments, rather than an average concentration for the whole impoundment. Concentrations in these animals may differ significantly with changes in PCB soil concentrations and soil types, both of which may have considerable small-scale spatial heterogeneity. If the data are extrapolated to other impoundments, this potential for spatial differences should be identified as an uncertainty.

4. *Comparison of egg concentrations for multiple avian species to egg-based TRVs.*

Panel's Response: Although the MSU studies have substantially increased the number of measured egg concentrations available for comparison to egg-based TRVs, the TRVs themselves are all literature-derived and of uncertain applicability to species present at the site. Hence, increasing the size of the egg concentration dataset may not significantly reduce the overall uncertainty in the egg-based HQs.

5. *Use of great horned owl pellet analysis and passerine nestling dietary composition analysis to refine estimate of dietary composition for purpose of dose modeling.*

Panel's Response: The owl pellet analyses and passerine nestling dietary composition data collected by MSU could provide more realistic dose models; however, it will still be necessary to account for potential variations in diet both between impoundments and between present and potential future conditions. In addition, it must be demonstrated that sample sizes from the MSU studies are sufficient to support their use in dose modeling.

1b. Effects of differing dietary preferences on extrapolating from the results of the MSU studies to other species. As an example, how may species-specific dietary preferences of the wrens or bluebirds evaluated in the MSU studies affect extrapolation of risk from these species to robins?

Panel's Response -- bluebird/house wren exposure (dietary dose) to other species (especially robin) exposure: Wren and bluebird diets cannot be directly extrapolated to other species; however, PCB concentrations in invertebrates and earthworms can be used to estimate dietary doses to other species, using information on dietary preferences of those species.

The Panel suggests that bluebird data may provide a bounding condition. This would help set the limits on the use of these data. Examination and contrast of feeding habits and – if available – information on sensitivity for other birds of this body size and metabolism to PCBs might lead to a conclusion that one cannot extrapolate directly but may be able to make inferences. By "extrapolation" the Panel means quantitative estimation of doses to one species using measured doses for another species, combined with models that quantify between-species differences in dietary preferences and intake rates. By "inference" the Panel means qualitative extrapolation of doses between species, using expert judgment rather than quantitative models. Supporting Reference: Blankenship et al. (2005) reports PCB concentration data for a variety of food items present in the Trowbridge impoundment.

1c. Potential effects of future conditions on risk, such as possible changes in habitat over time due to natural succession or anthropogenic changes to enhance recreational use. Some examples include lowered water table and reduced soil moisture content related to dam removal, transition to meadows including short grass habitat or succession to mature hardwood forest.

Panel's Response – Time-related extrapolation issues: Future land use, natural succession, and reduced inundation frequency can all be expected to change the utilization of the formerly inundated areas by key receptor species, and possibly also change the diets of those species that utilize these areas. For this reason, conclusions concerning risks to bluebirds, house wrens, robins, and great horned owls reached in the papers published by the MSU group may not hold in the future. However, provided that the influences of future land use, succession, and inundation on habitat suitability and prey availability can be predicted, it may still be possible to use the MSU data to predict foraging, diet composition, and PCB exposures under future conditions. It also is instructive to note that the bluebird study identified flooded habitat in the Trowbridge impoundment (as compared to the Fort Custer reference area) as one potential reason for reduced productivity. As succession-related changes occur, or as changes in the hydrologic regime follow from removal of remnant dam structures, habitat may improve for these and other passerines, thus resulting in higher productivity.

To facilitate consideration of the effects of natural succession, existing information on succession patterns in the riparian habitat that borders river systems in Michigan should be used to define two to four general succession conditions and associated biota. Habitat characteristics associated with each of these stages could then be used to predict the diets of passerine birds and great horned owls present in each impoundments as functions of succession stage.

Panel's Response – Evaluating Future Ecological Risks: The former Kalamazoo impoundments will likely undergo some degree of ecological succession over the next several decades. The rate of change will probably depend on topography, hydrologic regime, associated soil moisture, and proximity to colonizing plant species. Future wildlife species that inhabit and/or forage within the former impoundments could be different from those that currently utilize these areas. Therefore, some consideration should be given to the degree to which the MSU studies could be used to consider risks to future communities.

There is little doubt in the scientific community that global climate change is underway and will continue for the next several decades, as documented most recently in the Intergovernmental Panel on Climate Change (2007) reports. While precipitation in the Great Lakes region may increase, temperatures and evapotranspiration rates are also expected to increase, with a net lowering of water levels being plausible, although uncertainty exists, especially about precipitation forecasts. As a result, climate change mitigation strategies involve anticipating increased climatic variability. In the present case, that means expecting that the present flood frequency is likely to change, perhaps increasing the frequency of inundation but also perhaps decreasing it.

Changes in habitats within the former impoundments will track, to a large degree, with changes in soil moisture and periodic inundation. Insights into possible future habitats within the former impoundments can be drawn from information on the types of plant habitats that are present within the Kalamazoo watershed. The Michigan Department of Natural Resources (2002) has identified seven major types of native plant communities in the watershed, related to a gradient of soil moisture and inundation. These are listed from the driest to the wettest:

1. Southern Hardwood Forest: Forests of dry upland sites with bur oak, black oak, or white oak dominating.

2. Mesic Southern Hardwood Forest: Forests that occur in moist soils and are dominated by beech and sugar maple.

3. Wet Lowland Forest: Forests characterized by willow or cottonwood, or silver maple or ash.

4. Sphagnum Bogs: Open, treeless wet areas dominated by heath-like shrubs and sphagnum moss.

5. Grassland-Savanna Complex: Includes the combination of prairie, sedge meadow, and savanna communities, characterized as treeless or with scattered trees and dominated by grasses or sedges, either wet or dry.

6. Marshes and Emergent Aquatic Communities: Treeless areas in which the water table is above the soil surface during most of the growing season.

7. Submerged Aquatic Communities: Dominant plant species are below or on the water surface. These communities are essentially lakes and ponds.

Based on the Panel's field observations and the topography of the former impoundments, it is possible that future plant-based habitats could include Categories 2 through 6. As soils dry, increased areal extent of forests could be expected (mesic southern hardwood and wet lowland). Along the spectrum of soil moisture and inundation, mesic southern forest habitat is probably most different from that that existed during the MSU study.

Faunal characteristics of mesic southern hardwood forests are described by MDNR (2002) and MSUE (2004). Area-sensitive birds that depend on these forests include forest warblers, flycatchers, thrushes, vireos, woodpeckers, and woodland raptors. The woodcock, a species that feeds on soil invertebrates and that is often used for evaluating ecological risks, is also found in these forests. Raptor species higher in the food web include Cooper's Hawk (*Accipiter cooperii*), Northern Goshawk (*Accipiter gentilis*), and Red-shouldered Hawk (*Buteo lineatus*). This forest system provides summer nesting habitat for many neotropical migrants, especially interior forest obligates such as black-throated green warbler (*Dendroica virens*), scarlet tanager (*Piranga olivacea*), and ovenbird (*Seiurus aurocappilus*). Rare songbirds of mesic southern forest include cerulean warbler (*Dendroica cerulea*, state special concern), prothonotary warbler (*Protonotaria* *citrea*, state special concern), Louisiana waterthrush (*Seiurus motacilla*), and hooded warbler (*Wilsonia citrina*).

Mammals include woodland vole (*Microtus pinetorum*, state special concern), red bat (*Lasiurus borealis*), hoary bat (*Lasiurus cinereus*), northern bat or northern myotis (*Myotis septentrionalis*), Indiana bat or Indiana myotis (*Myotis sodalis*), eastern pipistrelle (*Pipistrellus subflavus*), least weasel (*Mustela nivalis*), and southern bog lemming (*Synaptomys cooperi*).

Temporary pools within mesic southern forest provide crucial habitat for reptiles and amphibians. Amphibian species most dependent on ephemeral pools in Michigan are spotted salamander (*Ambystoma maculatum*), blue-spotted salamander (*Ambystoma laterale*), chorus frog (*Psuedacris triseriata*), wood frog (*Rana sylvatica*), gray tree frog (*Hyla versicolor*), and American toad (*Bufo americanus*). Rare herptiles associated with these pools include small-mouthed salamander (*Ambystoma texanum*, state endangered), marbled salamander (*Ambystoma opacum*, state threatened), Blanding's turtle (*Emys blandingii*, state special concern), and copperbelly water snake (*Nerodia erythrogaster neglecta*, state endangered). Reptiles associated with mesic southern forest include black rat snake (*Elaphe obsoleta obsolete*, state special concern) and eastern box turtle (*Terrapene carolina carolina*, state special concern).

With respect to future remedial decisions, the Panel thinks that it is important to consider the trophic connection between PCBs in floodplain soils and receptors. To the degree that this can be focused on a spatially explicit and trophic-level basis will lower the uncertainty in assessment and decisions. For this reason, the datasets for soils, earthworms, and small mammals are particularly valuable for constructing the base of food webs that link PCBs in soils to animals that feed on soil invertebrates and on animals that feed higher in the trophic level (i.e., they feed on the small mammals and birds that rely on soil invertebrates as part of their diet).

Shrews: MSU obtained data for shrews, voles, various passerine birds, mink, and owls. Shrews are recognized as animals that forage on a wide variety of soil invertebrates including earthworms. The Panel has indicated elsewhere that these small mammals are useful to consider as ecological receptors and as prey for various bird and mammal species. There are several shrew species and they can occur in a wide variety of habitats. For example, the short-tailed shrew (*Blarina brevicauda*) inhabits inhabit a wide variety of habitats and is common in areas with abundant vegetative cover; they favor cool, moist habitats because of their high metabolic and water-loss rates (USEPA 1993). Studies conducted in the Housatonic River Watershed showed that the species occurs in a wide variety of habitats, including hardwood and mesic forests (USEPA 2004). Therefore, shrews can likely serve as appropriate ecological receptors across a wide variety of habitats that might develop within the former impoundments.

Voles: The MSU study collected data on and evaluated risk to the meadow vole. The meadow vole inhabits grassy fields, marshes, and bogs; compared with the prairie vole, the meadow vole prefers fields with more grass, more cover, and fewer woody plants (USEPA 2004). Therefore this vole would not be likely to occur in forest systems. Other voles and mice would be present.

Some consideration would need to be given to how well the information for meadow vole could be applied to other small mammals with food habits similar to the meadow vole.

Mink/Weasels: The MSU study considered the mink, which is expected to inhabit floodplains bordering rivers and streams. Minks are found associated with aquatic habitats of all kinds, including waterways such as rivers, streams, lakes, and ditches, as well as swamps, marshes, and backwater areas; they prefer irregular shorelines to more open, exposed banks as well as brushy or wooded cover adjacent to the water, where cover for prey is abundant and where downfall and debris provide den sites (USEPA 2004). Therefore, the drier the habitat, the less likely it will be utilized by mink. However, the least weasel is found in mesic southern hardwood forests and could represent a weasel species that might utilize these areas for habitat and food. The species feeds primarily on rodents and thus could be a receptor to consider in forest habitats. The long-tail weasel is another species that can live in a variety of habitats.

Higher Trophic-Level Mammals (e.g., Weasels): The Panel recognizes that evaluation of risks to mink is confounded by the fact that this species makes use of aquatic as well as terrestrial environments as a source of food. However, when considering potential future risks of PCBs in current floodplain soils, USEPA and the PRPs may want to consider a mammal species that feeds on small mammals and birds. As environments become progressively drier, this exposure route would provide a linkage between exposure of higher trophic levels to PCBs in soils. The Panel offers no recommendation on what this species might be, but suggest it be considered as a part of developing a comprehensive conceptual model. The Panel also notes that the least weasel is found in mesic southern hardwood forests and could represent a species that might utilize these areas for habitat and food. The species occurs infrequently and for that reason may not be the best representative. However, it does feed primarily on rodents and thus could be a receptor to consider in forest habitats. The long-tail weasel is another species that can live in a variety of habitats. The Panel mentions these species to the extent that they or other higher-trophic level mammals might be considered for evaluating future risks for floodplains.

Woodcock: This bird species was not considered in the MSU study, but this species could be an appropriate receptor for future habitats. Woodcock inhabit both woodlands and abandoned fields, particularly those with rich and moderately-to-poorly drained loamy soils, which tend to support abundant earthworm populations (USEPA 2004). In the spring, males use early successional open areas and woods openings, interspersed with low brush and grassy vegetation, for singing displays at dawn and dusk. Females nest in brushy areas of secondary growth woodlands near their feeding areas, often near the edge of the woodland or near a break in the forest canopy. During the summer, both sexes use second growth hardwood or early successional mixed hardwood and conifer woodlands for diurnal cover.

Panel's Response – Implications for the Earthworm Pathway: The BERA indicated that the relatively high concentrations of PCBs in earthworms make that the critical pathway for exposures in the Kalamazoo River Basin, showing up in the exposure pathway through the robin. However, the MSU studies did not emphasize this critical pathway and may not have sufficient data to reflect the potential risk to wildlife through earthworms. MSU did not collect data on the robin, in part because few earthworms were found in the former impoundment areas, presumably because of relatively frequent inundation by river water. As noted elsewhere, this condition may not remain the case in future scenarios. Even if robins are not currently foraging heavily on-site (because of availability of better foraging habitat in adjacent uplands), small mammals such as shrews are present on site and would be expected to feed on earthworms. Although the former impoundment areas of concern do not seem to be good robin habitat at present, consideration of the soil—earthworm exposure pathway is essential for evaluating risks to other earthworm-eating wildlife such as woodcock, snipe, shrews, and robins.

For example, in a risk assessment of a variety of soil contaminants at Oak Ridge National Laboratories (ORNL) (Efroymson et al. 1997), soil cleanup goals for PCBs were driven by shrews and then woodcocks, leading to values of 0.371 and 0.655 mg•kg^{-1} soils, respectively, based on a hazard quotient (HQ) model similar to the models used in the Kalamazoo River BERA. Note that the ORNL PCB cleanup goals are significantly below current soil PCB concentrations in the former impoundments on the Kalamazoo River. On the other hand, preliminary cleanup goals for floodplain soil at the Housatonic River site are far higher (21.1 mg•kg^{-1} to 43.5 mg•kg^{-1}) based on results of a site-specific study of short-tailed shrew population dynamics (USEPA 2004, GE 2005).

Given the wide range of cleanup values derived for other sites with PCB-contaminated soil, a focus on shrews would seem to be particularly appropriate to address the potential earthworm pathway exposures. MSU's October 05, 2001 update report (MSU 2001) provides data on tissue PCB concentrations in shrews collected from the former Trowbridge impoundment. Table 4-5 of that report summarizes tissue concentration data for 17 shrews. These data could be used to estimate doses to predators such as foxes and great horned owls that would be expected to prey on shrews. They could also be used to directly address risks to shrews; however, the need to rely on TRVs derived from rat studies would make any such assessments uncertain. Two plausible future scenarios exist that would likely change the frequency of flooding of these areas, thereby providing much more suitable habitat for earthworms, and consequently increasing the potential PCB exposures to earthworm-eating birds and mammals:

- The representative from the Michigan Department of Environmental Quality (MDEQ) stated that there remains a firm commitment by the State to complete the removal of the remnant dam structures, thereby allowing a free-flowing Kalamazoo River. It is apparent that by removing these structures, the River water levels would be lowered by a few feet, particularly in the lower half of the former Trowbridge impoundment upon removal of the remnants of the Trowbridge Dam. In that circumstance, the frequency of inundation of the former impoundment sediments would decrease significantly, likely to the degree that an earthworm population would become established.

- As mentioned earlier there is little doubt in the scientific community that global climate change is underway and will continue for the next several decades. While precipitation in the Great Lakes region may increase, temperatures and evapotranspiration rates are also expected to increase, with a net lowering of water levels being plausible, although uncertainty exists, especially about precipitation forecasts. As a result, climate change mitigation strategies involve anticipating increased climatic variability. In the present case, that means expecting that the present flood frequency is likely to change, perhaps increasing the frequency of inundation but also perhaps decreasing it.

Panel's Response – Effects of Weathering of PCB mixture on Toxicity Issue: Commercial PCB formulations (e.g., Aroclors) contain complex mixtures of many congeners. Once released into the environment, the relative concentrations of congeners in various environmental matrices change based on differences in volatilization, abiotic degradation, adsorption to soil/sediment, biotic metabolism, etc. The BERA emphasized total PCBs when calculating ecological risks because PCBs are regulated on a total and not congener-specific basis in Michigan (page 3-1). Even so, the BERA recognized the potential for weathering of PCB mixtures in the environment (page 4-7) and the relatively greater persistence of PCBs with five or more chlorine atoms (page 4-3). Furthermore, the section on derivation of TRVs for the BERA includes a discussion of both Aroclor-based and TEQ-based risk assessments (Appendix D pp. 26-31). This section of the BERA cites two previous papers from the MSU laboratory (Ludwig et al. 1996, Giesy and Kannan 1998) which concluded that "in general, risk assessments based on the original source of Aroclor are likely to underestimate the risk of bioaccumulated PCBs." So while risk estimates and remediation goals may be expressed in terms of total PCBs on account of regulatory requirements, MSU's congener-specific data is a major **strength** that allows the examination of congener patterns, total TEQs, and the contribution of individual congeners to total TEQs. This congener-specific approach reflects the current state of the science in this field.

MSU's conclusion about the weathering of PCB mixtures and reduction in the relative potency of the congener mixtures in Kalamazoo River floodplains (Blankenship et al. 2005) deserves careful examination, especially since this same laboratory has argued that differential weathering, metabolism, and/or bioaccumulation has led to enrichment of toxicity in higher trophic levels of Great Lakes food webs (Giesy et al. 1994a and b, Ludwig et al. 1996 provide several good summaries and reviews; many other papers from the MSU lab also could be cited). The specific arguments using relative potency (RP) factors to demonstrate weathering and reduction of toxicity (Blankenship et al. 2005) are weak. For instance, when examining the great horned owl (Fig. 4 and associated discussion), Blankenship et al. (2005) concluded that RPs show a reduction in toxicity for trophic transfers between small mammals and shrews to owls. Blankenship et al. (2005) downplay the RP of 1.9 for transfer from robins to owls because "it is unclear how much of the great horned owl diet may consist of robins and similar passerine species" (p. 5960). However, Strause et al. (2008) reported that passerines constituted a significant proportion (22%) of the owls' dietary mass at Trowbridge. So in fact there is evidence for enrichment of potency along this important trophic pathway. Furthermore, the two pathways with low RPs in Blankenship et al. (2005) are shown by Strause et al. (2008) to be negligible or irrelevant because they account for only small fractions of the owls' dietary mass at

Trowbridge (0.2% for shrews and 6% for other small mammals). RPs for these pathways are meaningless if owls consume few of these food items.

The assertion that BMFs cannot be applied directly to TEQs (Blankenship et al. 2005: 5959) is also questionable. The citation of only a single study to support this argument does not reflect the breadth and depth of published studies on TEQs in food webs. Furthermore, BMF considerations for total PCBs and TEQs are extremely similar from chemical and mathematical perspectives. Mathematically, total PCBs are calculated as the sum of individual congener concentrations with a relative weighting of 1 for each congener. TEQs are calculated as a sum using TEFs as weighting factors. Any changes in relative concentrations of individual congeners (i.e., preferential weathering, metabolism, or accumulation of particular congeners) will influence concentrations (and BMFs) of both total PCBs and TEQs. While TEQ BMFs alone cannot be used for inter-site comparisons, examination of these BMFs in conjunction with congener patterns and the fractional contributions of congeners to total TEQs provides a wealth of information about bioaccumulation patterns and Ah-receptor toxicity at individual sites and for comparisons of multiple sites.

Panel's Response – Higher PCBs in House Wrens: The high concentration of TEQs (both absolute and relative to total PCBs) in house wren eggs, nestlings, and adults is a significant finding and has potential implications for the avian exposure and productivity studies. These elevated concentrations indicate significant exposure to and/or preferential storage of toxic congeners in this species (Table 3.1). Clearly bioaccumulation/retention patterns of toxic congeners vary significantly amongst only the three avian species that were assessed. When considering the diversity of avian species in the Kalamazoo River floodplain, other birds with preferential accumulation and/or retention of planar PCBs may also vary significantly. (Note: the high TEQ concentrations in house wrens are not sufficiently discussed in the MSU papers.)

Table 3.1. Relationship between avian TEQs and total PCBs in birds at the former Trowbridge impoundment

Species/Tissue	Mean avian TEQs (ng/kg)	Mean total PCBs (mg/kg)	Ratio of TEQs/PCBs
House wren eggs	423	8.2	51.6
House wren nestlings	89	1.4	63.6
House wren adults	107	3.2	33.4
Bluebird eggs	57	7.4	7.7
Bluebird nestlings	6.7	1.7	3.9
Robin adults	3.9	0.92	4.2
Great horned owl plasma	0.69	0.49	1.4
Great horned owl eggs	13	7.2	1.8

Data from Blankenship et al. 2005

3.2.2 Question 2: What are the relative strengths, limitations, and uncertainties associated with the productivity assessments conducted by MSU on passerines and great horned owls (Neigh et al. 2006b, 2007; Strause et al. 2007a, 2008).

Panel's Response: Studies of productivity of the bluebirds, wrens, and great horned owls provide useful, qualitative evidence of reproductive performance of on-site species. The **strength** of these studies is that they are directly measuring one of the assessment endpoints ("do PCBs affect reproduction of birds?"). Field measurements are generally preferable to laboratory-based studies as they include much greater realism, including the fact that contaminant-induced changes are not always additive to other stressors that can reduce productivity (weather, predators, etc.). Of course, this is a **limitation** in field studies as well, as it requires a very large sample size to be able to apportion causality to observed effect and to statistically show differences among local populations. The **limitations** of the study are: the small samples sizes (which in some cases are insufficient to draw defensible conclusions); issues with pseudo-replication and other aspects of study design (also see Appendix A for further comments on pseudo-replication); reliance on aquatic organisms for a portion of the diet of the bluebirds and owls; lack of accounting for observational artifacts (such as time of nest initiation or failure) with the great horned owl study; the confounding effects of habitat differences between the KRSS sites and the reference site (Fort Custer); the situation in which the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge (box use is known to be significantly affected by familiarity of the birds with the placement of the boxes); and spatial mismatch between the size of the great horned owls' home range and the size of the study areas (with only a few nesting territories present in each area, relatively short-term studies offer limited representativeness of the ecological, reproductive, and exposure-related variables affecting this species). The full value of MSU's passerine reproductive data is uncertain until a more thorough reanalysis is conducted. (see below).

Supplemental Issues to Consider in Question 2

> ***2a. Strengths and limitations of directly measuring productivity in the field compared to extrapolating from controlled laboratory studies.***

Panel's Response: The MSU studies did not characterize population effects because productivity measures alone, without information on survival rates, cannot predict population consequences. In this regard, the assertion that one "bad year" and two "good years" is evidence that contaminants are not affecting productivity is not supportable. There is evidence from other studies that interactions of other environmental stressors (e.g., climate, parasites) with contaminants can result in reduced productivity, while the same contaminants in otherwise "good years" will not (Nagy, Schumaker, Fairbrother et al., unpublished data on western bluebirds). Furthermore, apportioning cause of nest failure is difficult, but methods now are available to provide a more quantitative estimate (e.g., Etterson et al. 2007).

Panel's Response – Great Horned Owl: The great horned owl is an appropriate receptor species for assessing the risks of PCBs to top predators consuming a mixture of foods from both terrestrial and aquatic food chains. (Note that other top predators exist in the Kalamazoo River floodplain, including red-tailed hawks, other raptors, snakes, and raccoons [the latter two predated bluebirds and house wrens in their nest boxes]). The following analysis does not specifically address extrapolations to other predators. The BERA concluded that the great horned owl was at significant risk to effects of PCBs accumulated through the terrestrial food web. The published MSU papers outline a rationale for monitoring great horned owls in field studies related to ecological risk assessments (Strause et al. 2007a, b, 2008, Zwiernik et al. 2007). However, the amount of data and strength of conclusions in the MSU studies were limited both by characteristics inherent to the biology of this species and by particular aspects related to how the studies were conducted or reported. These **limitations** reduce the ability of the owl dataset to assess risks to top predators and inform future risk management activities.

Principles of ecosystem energetics dictate that the amount of energy per unit area available to top predators is relatively low, and hence they are constrained to have large home ranges and low population densities. This limits the number of owls that can be supported by the formerly impounded/floodplain areas along the Kalamazoo River, leading to small sample sizes in the MSU studies. Furthermore, not all variables were (or could be) measured at all nests/areas, further reducing sample sizes for some variables, in particular reproductive productivity and call-count indices (see Table 3 of Strause et al. 2007a and Table 1 of Zwiernik et al. 2007 for sampling summaries). Low sample sizes and lack of replication reduced the statistical precision and power for analysis of many variables and eliminated the possibility of inferential statistics for productivity.

Further statistical questions are raised because of uncertainties in the temporal—spatial distribution of sampling and the independence of the sampled nests. None of the published papers shows or describes the spatial distribution of sampled nests within a given year or between years. Figure 1.2 in Giesy and Zwiernik (2008) provides a map of owl nest platforms and egg collections at Trowbridge for the all years of the study, but the nests that were occupied and sampled are not designated. Five owl eggs from Trowbridge were collected for residue analysis, but Figure 1.2 shows only 3 egg collection locations. Presumably several eggs were collected from the same nest or territory during different years, but the specifics are not clear. Beyond eggs, presumably some of the same pairs/territories were monitored in multiple years, and hence data would not necessarily be independent. Again, these relationships cannot be determined from the papers or presentation.

The assessment of owl reproductive productivity (fledglings per nest) was particularly limited by small sample sizes and lack of replication of sites; thus the published studies over-interpret the strength of this data set. Comparisons of great horned owl populations between Trowbridge and Fort Custer sites cannot be done in any meaningful way because of small sample sizes. Only one active nest was monitored for productivity at the reference site and six at Trowbridge. Several methods have been put forward for calculating nesting and fledging success to account for problems associated with timing of observations, the most common of which is the Mayfield (1975) method. Owl nest productivity might be recalculated using this alternative method to see if additional information about egg production, and hatching or fledging success can be ascertained. However, the data set might be too small, in terms of both sample sizes and the types of collected data, for this reanalysis to be done. Clearly Fort Custer was an insufficient reference site for productivity studies. MSU incorrectly concludes (Strause et al. 2007a, Zwiernik et al. 2007) that the “mean” (n=1 for reference) productivity of 1 young/active nest was not “significantly” different between Trowbridge and the reference site.

Furthermore, both of these MSU studies (Strause et al. 2007a, Zwiernik et al. 2007) conclude that Trowbridge productivity was consistent with the productivity found by Holt (1996) in a multi-decade study of 906 great horned owl nets surrounding Cincinnati, Ohio (Holt 1996). This comparison to only *one* other published study on great horned owl productivity is a very limited ecological analysis, and additional studies should be considered. GHO productivity can vary greatly depending on ecological factors, especially food supply. When major prey items such as hares are particularly abundant, productivity can be as high as 2.5-2.6 fledglings per successful nest (Houston 1987 and Houston and Francis 1995, as cited in Holt 1996). Hence, there is little support for the conclusion that a productivity of 1 fledgling per active nest is in fact “normal” for southwestern Michigan or the upper Midwest.

2b. Extrapolation of results from site-specific productivity studies to other species such as the robin, which was the receptor species considered in the Final Baseline ERA (CDM, 2003).

Panel's Response: Site-specific studies of species productivity can be extrapolated to other species but only with added uncertainty. Species differences in diets and home range locations would suggest that exposures would differ and, therefore, effects may as well. Compensatory / depensatory factors may have some similarities (e.g., weather – exceptional cold just as young are hatching) or not (e.g., predation; ability to re-nest; size of clutch). On the other hand, species that are in similar feeding guilds (e.g., primarily insectivorous) and of the same type of life history strategy may be sufficiently similar in exposure and interactions with their environment to allow extrapolation of effects from one to the other. However, none of the species studied by MSU (eastern bluebird, tree swallow, house wren, or great horned owl) has diets that are entirely similar to that of the robin. Furthermore, the tree swallow may be particularly insensitive to PCB effects (see below; section 2c), and so would not be a good model for other passerines. Extrapolation of bluebird productivity to robins would be confounded by the fact that bluebirds eat a significant amount of aquatic invertebrates. House wren diets are entirely terrestrial invertebrates but generally do not include earthworms or fruit, both of which are consumed by robins during certain seasons. House wrens have a clutch size of 3-7, while robin clutches are smaller (3-4), and male house wrens tend to be more promiscuous than male robins, which may increase productivity per unit area (http://www.birds.cornell.edu). All of these attributes could be accounted for in a qualitative uncertainty discussion if bluebird or house wren productivity measures were to be extrapolated to robins. Given the difficulties with sample size and study design, the added uncertainty reduces the usefulness of such extrapolations.

Panel's Response – Passerines Productivity Methods: Many previous field studies have shown that reproductive and developmental endpoints are useful for assessing risks of PCBs, dioxins, and furans to a variety of avian species, including colonial water birds (gulls, terns, cormorants, and herons), raptors (especially eagles), and passerines (especially tree swallows) Such field studies generally complement the wealth of laboratory studies on the reproductive effects of these chemicals in birds. Hence, the reproductive productivity studies of Kalamazoo River passerines were well founded and provided significant real-world biological data that were, for the most part, absent from the BERA. However, MSU's productivity studies were limited by several factors, including (perhaps) ecological factors beyond the control of the investigators (e.g., low sample sizes for bluebirds at Trowbridge) but also by issues related to the study design, data analysis, and clarity of data reporting. The use of hypothesis testing is not appropriate with the collected data (see Appendix A). The generally low sample sizes limit the applicability of the conclusions that can be drawn. In particular, the risk of a false negative conclusion (type II error) is high with small sample sizes and the widths of confidence intervals on parameters are too wide to be useful in making practical decisions. Despite the small sample sizes, the data contain a number of indications that reproductive success was lower at Trowbridge compared to the Fort Custer reference area. However, inferences about causation by

PCBs are unsupported and would require replication of study areas and/or additional exposure-response models. Additional analyses might include an analysis of ecological covariates (e.g., habitat characteristics, weather conditions) by building statistical models. Several important spatial-temporal issues related to the study design and data analysis are not reported adequately in the SOPs and publications. (Note: As with the great horned owl studies, the passerine SOPs [260, 262, and 264] seem to reflect initial plans and were not updated to reflect the methods that were actually used.) There appears to be no SOPs for bluebirds or house wrens.

Figure 1.2 in Giesy and Zwiernik (2008) provides a map of nest boxes and egg collections at Trowbridge for the all years of the study, but the nests that were occupied and sampled are not designated. None of the published papers or SOPs shows or describes the spatial distribution of sampled nests within a given year or between years. These spatial-temporal relationships are of interest for several reasons. The Stage 1 Assessment for the Kalamazoo River NRDA concludes that different areas of the former impoundments present different levels of risk to terrestrial passerines (see Fig. 7.19 in Stratus 2005). The MSU studies potentially provide the opportunity to examine whether passerine reproduction was affected in the more-contaminated sections of the Trowbridge impoundment, but this cannot be done without more information on the location of active nests. More detailed information on the nests also would be useful for censoring and re-categorizing the data. During the May tour of the Trowbridge impoundment by the Panel, MSU scientists described instances in which bluebirds nested in flooded habitats, which would generally be considered atypical for this species and perhaps inappropriate for evaluating risks of PCBs in soils. However, these nests (or any other nests) cannot be identified and removed from the analysis based on the descriptions in the MSU publications. With respect to time, a more complete reporting of initial nests and re-nests would be beneficial. Given the generally equal sample sizes reported for early and late nests (Table 3 of Neigh et al. 2007), these categories seem to reflect an even split of nests rather than a more detailed classification of first and second nest attempts (e.g., based on documented first and second nest attempts in individual boxes).

Another methodological issue was the removal of eggs for chemical residue analysis. This practice introduces significant inaccuracy into the estimates of reproductive success. Consider a nest with 3 viable eggs and 1 nonviable egg. If no eggs are removed and the nest is monitored through hatch, the true hatching success of 75% and brood size of 3 would be known. However, random removal of 1 egg for chemical analysis would result in inaccurate estimates of hatching success. Removal of a viable egg (which would happen with 75% probability) would yield a hatching success of 67% and calculated brood size of 2.75. Removal of the nonviable egg (25% probability) would yield a hatching success of 100% and predicted brood size of 4. Similar considerations exist for numbers of fledglings. MSU's calculations of predicted brood size and predicted number of fledglings cannot fully remove this inaccuracy and variability, nor do they account for biological effects of artificially reduced clutch size (e.g., increased survival of young in manipulated nests because more food is available to each chick). The effects of this inaccuracy were disproportionately greater at Trowbridge. Eggs were removed from 40% of bluebird nests and 39% of house wren nets at Trowbridge, compared to 25% and 20% at Fort

Custer (based on sample sizes in Tables 1 and 3 of Neigh et al. 2007). The inconsistent removal rates may have impacted the comparison of hatching success among sites.

Interpretation of variables related to hatching and fledging is often complicated by the definition of these terms, which influences the particular individuals and nests that are included in the calculation of a given variables. Calculation of some of these variables in the MSU studies is not explained clearly. Furthermore, samples sizes often do not agree or add up between Tables 2 and 3 or within Table 3 of Neigh et al. (2007). However, based on the definitions of some of MSU's variables and the trend for declining sample sizes with the advancement of the reproductive cycle (from laying to fledging), it seems that nests that failed early were often ignored for later calculations. For example, fledging success included only "successful" nests in which at least one young fledged, ignoring the biologically relevant nests that failed to produce any young. The consistent decreases in sample sizes from clutch size to hatching success are unexplained. Hatching success should include nests in which no eggs hatched. The predicted number of fledglings is based on an undefined but small (in fact, the smallest) subset of nests. This variable should be calculated for all nests that were initiated. MSU's subdivision of the reproductive cycle into several segments is useful for determining what might be happening during different periods, but it also potentially obscures the greater question of overall reproductive success (i.e., the number of fledglings based on all initiated nests). Incremental effects during various subsections of the breeding cycle might be insignificant in and of themselves but might add up to a larger, more significant effect over the entire reproductive cycle.

To calculate an overall measure of reproductive success using MSU's data, the Panel multiplied clutch sizes (# of eggs laid/initiated nest) by the productivity (number of fledglings/egg laid) to give a fledging rate of number of fledglings per nest initiated (see Table 3.4 below). Fledging rates were 47% lower for bluebirds and 18% lower for house wrens at Trowbridge compared to Fort Custer. Note that removing the Trowbridge bluebird female that experienced repeated failures has little influence on this conclusion. Hence, overall reproductive success appears to have been much lower at Trowbridge, especially for bluebirds.

MSU's presentation of nest success (Table 1 and associated text in Neigh et al. 2007) is also parsed into smaller subsets (years and causes), perhaps missing larger picture differences. Simply summing the data across years (see Table 3.3 below) suggests much lower rates of nest success in bluebirds at Trowbridge and marginally higher rates of nest abandonment in both species at Trowbridge. These variables should be subjected to statistical analysis.

Table 3.2. Coefficients of variation based on Table 3 of Neigh et al. 2007.

	Fort Custer			**Trowbridge**		
	mean	**sd**	**cv**	**mean**	**sd**	**cv**
<u>Eastern Bluebird</u>						
hatching success	0.79	0.31	39	0.59	0.42	71
fledging success	0.96	0.21	22	0.83	0.35	42
productivity	0.76	0.34	45	0.47	0.44	94
clutch size	4.2	1	24	3.6	1.5	42
predicted brood size	3.8	1.1	29	3.3	1.4	42
predicted number of fledglings						
<u>House wren</u>						
hatching success	0.81	0.25	31	0.64	0.41	64
fledging success	0.92	0.34	37	1	0	0
productivity	0.74	0.3	41	0.64	0.41	64
clutch size	5.7	1.1	19	5.4	1.4	26
predicted brood size	5	1.6	32	4.6	2	43
predicted number of fledglings	4.8	1.6	33	4.6	2	43

Table 3.3 Fledging rates based on all active nests in which clutch size was measured (data from Table 3 of Neigh et al. 2007).

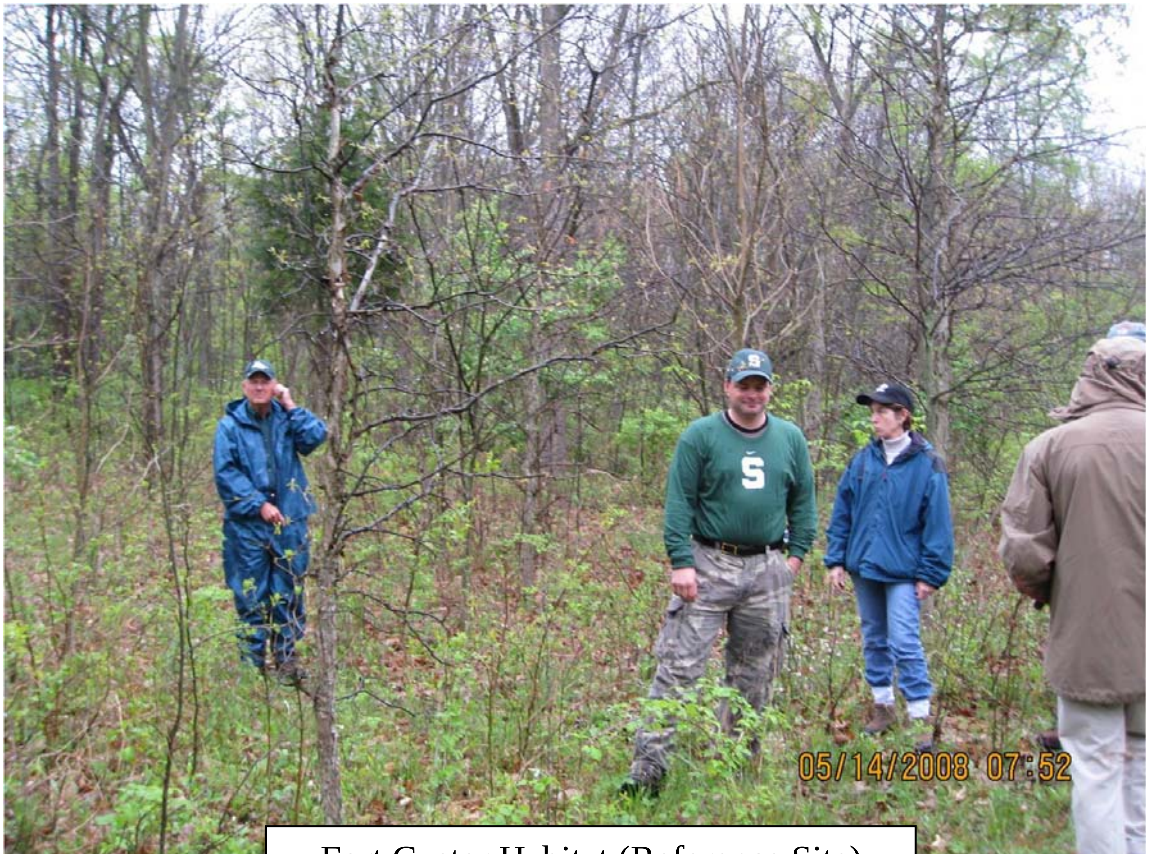
	Fort Custer	Trowbridge	Trowbridge % below Fort Custer	Trowbridge without female that failed repeatedly	Trowbridge % below Fort Custer without female that failed repeatedly
Eastern Bluebird					
Clutch Size (eggs/nest)	4.20	3.60	-14.3	3.60	-14.3
Productivity (# young/egg laid)	0.76	0.47	-38.2	0.51	-32.6
Fledging Rate (# young/nest)	3.19	1.69	-47.0	1.84	-42.2
House Wren					
Clutch Size (eggs/nest)	5.70	5.40	-5.3		
Productivity (# young/egg laid)	0.74	0.64	-13.5		
Fledging Rate (# young/nest)	4.22	3.46	-18.1		

Table 3.4. Nest fate for all years combined (data from Table 2 of Neigh et al. 2007)

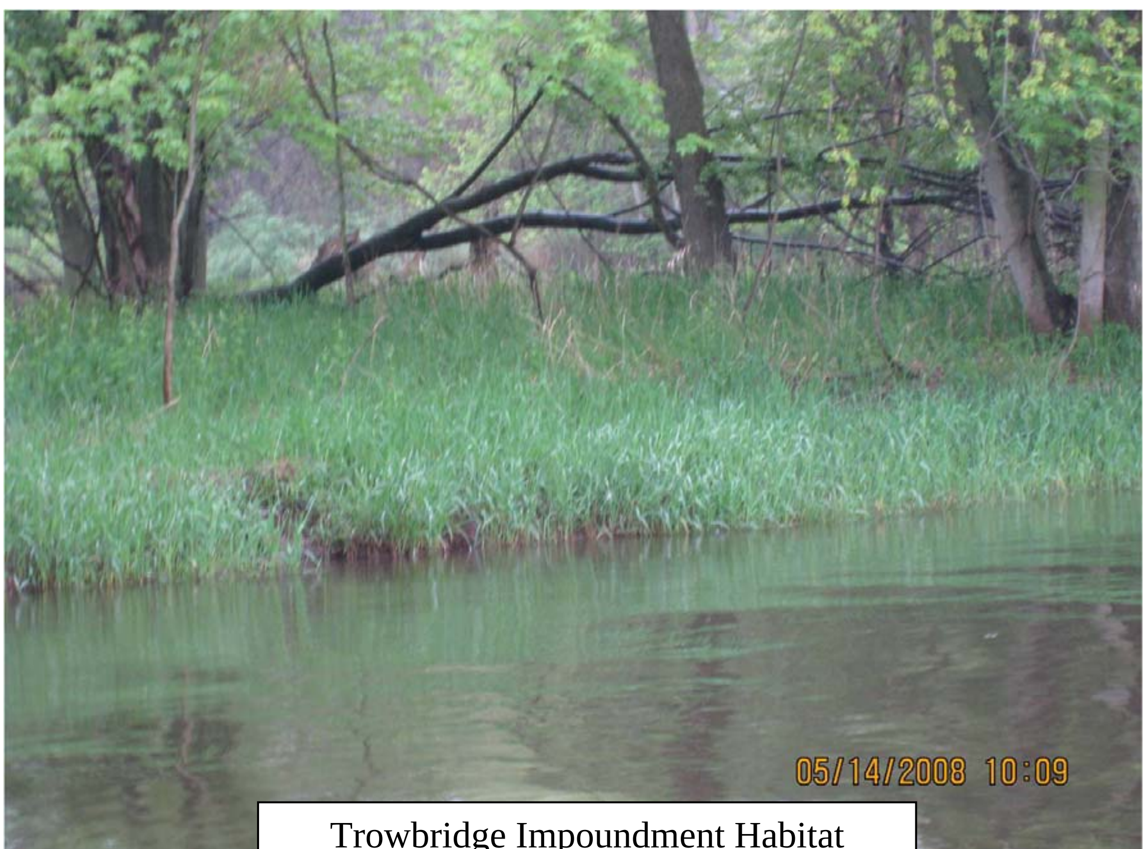
	Fort Custer total	% of Total	Trowbridge total	% of Total
<u>Eastern Bluebird</u>				
Successful	43	75.4	8	44.4
Abandoned	4	7.0	3	16.7
Total	57		18	
<u>House Wren</u>				
Successful	55	77.5	25	73.5
Abandoned	0	0.0	5	14.7
Total	71		34	

2c. Evaluation of potential causal factors (e.g. PCB concentrations, habitat differences, etc) associated with any difference in measures of productivity in passerines relative to the reference site.

Panel's Response: Comparison of avian productivity between Trowbridge and Fort Custer is difficult because of small samples sizes, issues with pseudo-replication (e.g., non-representative data) and other aspects of study design, lack of accounting for observational artifacts (such as time of nest initiation or failure) with the great horned owl study, the large effect of one bluebird female's nest failure on the overall success rate of the local population, and the confounding effects of habitat differences among the KRSS sites and Fort Custer. Habitat differences include more open areas at KRSS and more riparian woods at the reference site. Complicating the interpretation is that the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge; box use (and subsequent contribution to population productivity) is known to be significantly affected by familiarity of the birds with the placement of the boxes (i.e., may be lower at Trowbridge due to unfamiliarity with the boxes even though the boxes were erected one year before the studies were done). Given these difficulties, it is not possible to make statistical inferences based on hypothesis testing about productivity on KRSS and compared to Fort Custer. The fact that nest boxes were provided at both areas did not significantly mitigate these habitat differences because nesting cavities comprise only one aspect of songbird habitats.



Fort Custer Habitat (Reference Site)



Trowbridge Impoundment Habitat

3.2.3 Question 3: What are the relative strengths, limitations, and uncertainties associated with the hazard quotient calculations performed by MSU to evaluate potential risk to passerines, great horned owls, and shrews (Neigh et al. 2007; Strause et al. 2007a, 2008)

Panel's Response: Hazard quotient calculations consist of two components: a dose estimate and a TRV. MSU's HQs replace many of the modeled doses used in the BERA with site-specific estimated derived from field studies. The use of new site-specific data is an important **strength** of the MSU HQs. The TRV estimates used by MSU, in contrast, were based on the same suite of previously published studies used in the BERA. Any differences between the TRVs by MSU and the TRVs used in the BERA reflect a combination of differing scientific judgments and differing degrees of conservatism. The Panel can comment on the scientific aspects of MSU's TRV selection process, but defers to USEPA with regard to the appropriate degree of conservatism.

As shown in Tables 3.5 and 3.6 (below), the BERA and MSU approaches differ substantially with respect to both dose estimation and TRV selection. The BERA approach relies heavily on measured concentrations in abiotic media, extrapolated to doses using site-specific and literature-derived transfer factors. The MSU approach relies almost exclusively on PCB concentrations measured in food items. MSU's site-specific approach should provide more realistic estimates of PCB exposures for the purpose of risk assessment; however, transfer factors such as those used in the BERA would still be necessary to calculate remediation goals for soil and sediment.

For passerines, differences between BERA and MSU HQs are driven by the TRVs. There is not much difference between the diet-based MSU and BERA HQs. The TRVs used in the BERA are somewhat lower than those used by MSU, but interchanging the BERA and MSU TRVs changes the HQs by only a factor of 2-3, probably smaller than the uncertainty associated with the original BERA and MSU HQ values. The MSU egg-based TRVs are much higher than either the BERA or MSU diet-based TRVs. Both the NOAEL and LOAEL egg-based TRVs are derived from field studies of nest productivity (tree swallow and robin) rather than controlled laboratory studies. For the great horned owl, both assessments use the same TRVs. The differences are attributable to the large differences in the dose estimate used in the BERA as compared to the MSU studies. The principal contributor to the dose estimate calculated in the BERA (Table C-1) is the estimated concentration of PCBs in robins, which was calculated from earthworm data using a literature-derived biomagnification factor. The MSU dose estimate was based on measured concentrations of PCBs in adult robins.

Table 3.5 Comparison of assumptions and dose calculation methods – BERA vs. MSU

Assumption or method	BERA	MSU
Abiotic media concentrations	U95 confidence bound on arithmetic mean (or max value, if sample size insufficient)	Unknown (used only for incidental soil ingestion)
PCB concentrations in food items	Maximum measured tissue concentrations for earthworms; site-specific soil to plant transfer factors for plant tissues	Measured tissue concentrations in all food items (average?)
Dose estimation	Measured/estimated concentrations in food items + literature-derived diet and standard metabolic parameters from EPA exposure assessment handbook.	Measured concentrations in food items + site specific diet + standard metabolic parameters from EPA exposure assessment handbook.
TRV	Passerines: NOAEL and LOAEL based on chicken (0.4 mg/kg-d to 0.5 mg/kg-d) Great horned owl: NOAEL and LOAEL based on screech owl (0.41 mg/kg-d – 1.2 mg/kg-d)	Passerines: NOAEL and LOAEL based on pheasant (0.6 mg/kg-d to 1.8 mg/kg-d). Alternative NOAEL and LOAEL extrapolated from egg-based TRVs using biomagnification factors (1.9 mg/kg-d to 14.7 mg.kg-d) Great horned owl: NOAEL and LOAEL based on screech owl (0.41 mg/kg-d – 1.2 mg/kg-d)

Table 3.6 Hazard Quotient Calculations Based on MSU or BERA Doses and TRVs

		TRV		
				MSU
Species	**Dose**	*BERA*	*MSU*	*(egg-based)*
Robin	*BERA*	**1.8-2.3**	0.5-1.5	0.06-0.47
	MSU	N/A	N/A	N/A
Eastern Bluebird	*BERA*	N/A	N/A	N/A
	MSU	**1.0-1.3**	0.3-0.85	0.03-0.3
House wren	*BERA*	N/A	N/A	N/A
	MSU	0.26-0.33	0.07-0.2	0.009-0.07
Great horned owl	*BERA*	**1.8-2.3**	**1.8-2.3**	N/A
	MSU	0.05-0.14	0.05-0.14	N/A

The HQ estimates in Table 3.6 show that, depending on the choice of assumptions and TRV values, HQ values for all three species can be either greater than or less than 1.0. Detailed discussions of the supplemental issues related to HQ development and interpretation are provided below.

Supplemental Issues to Consider

3a. Choice of toxicity reference value (TRV), including relevance to receptor species and quality of study (e.g., duration, inclusion of sensitive life stages, exposure range, endpoints measured).

3b. Uncertainty resulting from extrapolating from laboratory study to field.

3c. Uncertainties in extrapolating from one species to another

Panel's Response – Toxicity Reference Values: This charge question addresses the appropriateness of MSU's toxicity reference values as well as the use of MSU's data to extrapolate to other species. The Panel is concerned about MSU's approach to developing TRVs for the PCBs and species of concern. The choice of TRVs is a very significant issue that drives risk assessment conclusions. In particular, MSU's TRVs used for calculating the hazard quotients (i.e., evaluating risks relative to field-measured or modeled exposures) are quite high and significantly influence their conclusions of no or minimal risk. Application of more protective (but realistic and accepted) TRVs may well lead to different conclusions (see Table 3.6). Given the critical importance of the TRV selection to the overall risk assessment and the resulting risk management decisions, the detailed rationale and basis used by MSU to derive the TRVs need to be articulated; the 27 July 2008 memo to the Panel did provide more information on the process used to derive TRVs, but concerns still remain.

Clearly, there should be developed a MSU technical document that describes all of the toxicity literature that was reviewed and the decision rules used to select studies for inclusion in the toxicity database. If a TRV is to be based on a single study, then the specific justification used to select the particular study to represent the effects level is needed, *including the rationale for not selecting alternative studies*. Similarly, if a TRV is derived from a statistical analysis based on several toxicity studies, the approach preferred by the Panel, then the basis for the derivation of the TRV must be explained and justified. The MSU memo in response to the Panel's request for an explanation of how TRVs were chosen indicated that a single study was selected for establishing each TRV value, but that leaves in question the validity of the TRV that was selected and, consequently, the basis for the risk analysis. It should the emphasized that use of an alternative TRV could lead to the opposite conclusions from those reached by the MSU Team, as illustrated in Table 3-6.

It also should be noted that the derivation of TRVs for PCBs in birds and mammals is not being done here for the first time. It would seem appropriate for there to be a thorough literature review of TRVs that have been used in other ecological risk assessments involving PCBs in riverine and riparian habitats to provide an indication of the range of avian and mammalian TRVs used in the published literature and other risk assessments. In general, toxicity values may vary greatly across PCB congeners, species, toxicity endpoints, test methodologies, etc., so selecting the particular study or studies to use for defining a TRV needs to be carefully considered, documented, and justified. Indeed, basing a TRV on a single toxicity study seems unwise because of these sources of variability and because the species of concern in the risk assessment are not species used in typical laboratory-based studies, i.e., there must be a reliance on cross-species extrapolations. Calabrese and Baldwin (1993) is a useful reference for TRV derivation methods.

The USEPA issued the final report on the EcoSSL (ecological soil-screening level) methodology last year, and although there is no Eco-SSL document for PCBs, the approach used in developing mammalian TRVs for polycyclic aromatic hydrocarbons (PAHs) would seem to be instructive here (see USEPA 1999, 2005, 2007) In that approach, USEPA examined a large database of published toxicity studies (involving several thousand studies) and applied objective criteria for selecting those particular studies that warranted inclusion in the EcoSSL database for PAHs (about 40 studies met the USEPA criteria). It would seem logical to follow a similar approach by developing explicit criteria to apply in selecting the particular toxicity studies to base TRVs upon. Once the studies were selected, EcoSSL provided a conservative (protective) methodology for deriving mammalian TRVs, including, among other things: looking only at population-relevant endpoints (mortality, growth, and reproduction) but not other endpoints (e.g., histological results); including all relevant data in a single assessment to characterize the toxicity variances; and under some conditions using a statistical metric from these data to derive the TRV value. See Figure 4.8 in USEPA (2005) for a detailed flowchart of the steps in derivation of wildlife TRVs; here, either a statistical approach is recommended or reliance on a single toxicity study, the latter if the highest-bounded NOAEL is below the lowest-bounded LOAEL for reproduction, growth, or mortality. (Note, however, that when USEPA followed this approach for PAHs in terrestrial mammals, a single toxicity test became the basis for the proposed screening TRV, but further examination of that particular study shows serious flaws in its experimental design [e.g., approximately 50% mortality in controls, and carcinogenesis as the endpoint instead of a toxicity], which seriously undermines the defensibility of that TRV. This further supports the Panel's recommendation not to base the TRV on a single study, because, in effect, any limitations in that selected single study may be transformed into constraints on the validity or defensibility of the risk assessment.)

USEPA also used a multiple-studies approach to setting water quality criteria (Stephan et al. 1985), in which the lowest 3 data points on the species sensitivity curve are selected and a triangular distribution fitted to them. Then the 5th percentile is calculated and used as the AWQC. Part of the rationale for using this multiple-study, statistical approach is to capture the potential range of variability in species toxicity responses to be protective of at least 95% of the species 95% of the time. Again, there would seem to be opportunity here to follow a similar approach, basing the TRV derivation as much as possible on USEPA-approved methods. If another process is followed, then there needs to be an explanation of the details of the process that was used and the rationale for following that approach, providing clear justification of the selected TRVs to the exclusion of other potential selections.

The information concerning TRVs that was provided in the publications about how TRVs were selected is too cryptic. For example, in the description by Neigh et al. (2006 ••• a or b ? •••) on how studies were qualified, there was no mention about congener specificity. Since it is known that PCB congeners differ significantly in their toxicity (and mode of action), it is extremely important that previous studies used to set threshold values be done with similar congeners or with toxicity-adjusted equivalents. For example, the ring-necked pheasant study by Dahlgren et al. (1972) used Aroclor 1254, and the initial source of PCBs at the Kalamazoo was Aroclor 1242 (which has since been weathered; see Blankenship et al. 2005). Yet no mention is made about which congeners are contained in these two Aroclors and which ones are likely to have more dioxin-like toxicity. Neigh et al. (2006) also used a study by Nosek et al. (1992) who did an IP injection with TCDD. Neigh et al. (2006) correctly identified the shortcomings of using this study for derivation of a TRV, in particular that the exposure route was via an IP injection. Many will argue that this is an unnatural exposure route and is not suitable for comparison with oral exposures. It would be instructive to have more justification by the MSU team for why this study is applicable Is it a “worst case” exposure, thus potentially providing a conservative risk estimate?

The Panel is also concerned about the use of tree swallow data for determining avian TRVs; this is questionable because tree swallows can accumulate high concentrations of PCBs yet show minimal or no heath/reproductive effects. There may be some taxonomic justification for extrapolating within passerines from tree swallows to bluebirds and house wrens. However, previous avian studies by the MSU lab have used much lower TRVs when considering risks to colonial water birds and bald eagles in the Great Lakes (Giesy et al. 1994 a. b). Considering the diversity of avian species in the floodplain habitats, some of these species are likely to be more sensitive to the effects of PCBs and therefore would be more appropriate for a conservative risk assessment. Note that using the egg avian TRVs for Aroclors 1242 and 1248 from the BERA assessment (Appendix D) in association with MSU’s exposure data would clearly produce HQs above one both for passerine species and owls.

The Panel suggests presenting all of the toxicity studies endpoints on the same graph (similar to that used by the EcoSSL or as was done by Jim Chapman for the BERA). If these studies are to be used for decision-making, there needs to be a thorough vetting and discussion to reach agreement by all parties on which TRV to use (or, perhaps, several TRVs to bound the probable effects range).

The Panel noted an inconsistency in the MSU TRV development approach in how the lowest-observed adverse effects level (LOAEL) values relate to the no-observed adverse effects level (NOAEL) values. NOAELs cannot always be determined from empirical data since the particular study might not have had a sufficiently low dose to have a result with no significant effects compared to the controls. In that case, NOAELs are occasionally derived from the LOAELs. That was the approach used by MSU, but a multiplicative factor of 3x was used for total PCBs, while a factor of 10x was used for TEQs. Either a rationale for the discrepancy needs to be provided, or else the same factor should be applied.

Neigh et al. (2006) estimated a NOAEL from a reported LOAEL in the pheasant feeding study by dividing the LOAEL by 3, based on an assumption that the reported LOAEL was “near the threshold for effects.” They also estimated a NOAEL from a LOAEL for the pheasant IP injection study (using TCDD; Nosek et al. 1992) by dividing by 10, as there were “”pronounced effects” occurring at the LOAEL dose level. This uncertainty should be accounted for by the risk managers when assessing the degree of confidence in the risk outcome. Overall, Neigh et al. (2006) attempted to be highly conservative in their choice of dietary TRVs. Nevertheless, prior to accepting their suggested dietary thresholds, the Panel recommends a thorough and detailed discussion with all parties on the strengths and limitation of the studies in the literature.

Strausse et al. (2007) developed a TRV for the great horned owl based on a study of dietary exposure of a screech owl. The **strengths** of this approach are the taxonomic similarity of the two species and the dietary route of exposure used in the controlled study. The **limitations** are that allometric dose scaling apparently was used (as per Sample et al. 1996), which has been shown to be inappropriate for chronic exposures (Sample and Arenal 1999; Luttik et al. 2005). However, the discussions of allometric adjustments to derive TRVs provided in Strause et al. (2008) for the GHO and in the MSU response to the Panel memo of 25 July 2008 for shrews do not provide details of the scaling factor used, and the language is somewhat ambiguous as to what adjustment was made. If MSU actually just normalized the dose values to a per-weight basis (which is *not* an allometric adjustment), then this is the correct approach to derive chronic TRVs. If, however, MSU actually did use an allometric scaling factor designed to address differences in metabolic rates across species, then this is not correct (and it would not be necessary, anyway, as the actual energetic needs of the GHO are directly known [Duke et al. 1973]). The resulting effect of the allometric approach on the TRV value cannot be determined without knowing what scaling factor was used.

Another **limitation** is that specific PCB congeners were not identified (i.e., to support an argument that the PCBs used in the screech owl study are of greater or similar potency to those at the KRSS). Again, a LOAEL was estimated from the NOAEL using an uncertainty of factor 3, which is always a subjective approach. The TEQ approach for threshold derivation used the TCDD study by Nosek et al. (1992) and followed the same approach as outline by Neigh et al. (2006).

In summary, the Panel strongly feels that since the selection of the TRV values is critical to determining the results of the HQ-based risk assessments, there must be adequate documentation and justification of the data and the process used to derive the TRVs. The use of a range of TRVs (and, consequently, HQs) in the risk assessment, each fully explained with respect to source and uncertainties, would enhance the utility of the risk analysis and support of the risk management process.

3.2.4 Question 4: What are the relative strengths, limitations, and associated uncertainties that should be considered when evaluating the results of these studies as potential lines of evidence in an area-specific risk assessment?

Panel's Response: The exposure data collected by MSU should be very useful for quantifying exposures of key ecological receptors addressed in area-specific risk assessments. However, for reasons discussed below, MSU's conclusions concerning risks to passerine birds and great horned owls may not be applicable.

Supplemental Issues to Consider:

4a. Study designs including (but not limited to) sample size, replication, temporal duration, and aggregation of data.

Panel's Response – Conceptual Model: The problem formulation stage of a risk assessment considers a number of important topics, including the specific purposes and scope of the assessment, the choice of receptor species, the identification of critical exposure pathways, the time frame under consideration, and the philosophy of protectiveness (i.e., the adoption of more or less conservative methods for choosing toxicity reference values, making extrapolations between species and locations, etc.). In the risk assessment paradigm, the resulting conceptual model greatly influences the design and execution of subsequent studies as well as the analysis and interpretation of data.

The conceptual models of the BERA and MSU studies differ in significant ways, both positively and negatively influencing the ability of the MSU studies to address questions posed by the BERA and validate and/or revise conclusions of the BERA. While the Panel has not been charged with reviewing the BERA, some comments regarding the BERA's conceptual model are helpful for comparison to MSU's conceptual model and assessment.

Overall the intent of the BERA appears to be broadly protective for a wide range of species during both the present and future time frames. Consistent with the conceptual model presented in the BERA, receptor species were selected for study and modeling based a number of criteria, including sensitivity and potential exposure to PCBs. A large number of studies were used to derive TRVs. Some TRVs were intentionally protective to account for the potential presence of sensitive species and extrapolation beyond the modeled receptor species.

By comparison to the BERA, the overall purposes, conceptual model, analyses, and conclusions of the MSU studies were more narrowly focused with respect to species and time and reflect a less protective approach. Further, MSU did not clearly articulate its conceptual model of the potentially important exposure pathways in the areas of concern, nor did they identify how the studies that were conducted contribute to the overall understanding of the ecological risks. (On the first point, the introduction of MSU's Sampling and Analysis Plan describes receptor species but does not delineate exposure pathways. Blankenship et al. 2005 models exposure pathways through a food web, but some of the specific food chains were shown to be insignificant by MSU's food preference data.) The absence of a comprehensive conceptual model of the ecological risks by the MSU team is a major **limitation** that leaves the results of their studies insufficient to challenge some of the conclusions drawn from the BERA.

One major purpose of the MSU studies was to compare conclusions resulting from multiple lines of evidence or different risk assessment approaches (e.g., top-down versus bottom-up approaches). (This objective was an organizing principle for 6 of the 8 published papers and was cited by Dr. Giesy at the May 13 charge meeting as one of two major purposes of their studies. The other major objective was to compare assessments based on total PCBs versus TEQs based on PCB congeners, which was the emphasis of one paper. The 8th paper emphasized other aspects of environmental chemistry in owls.) The MSU studies made this comparison of risk assessment methods considering only present conditions and only the few species selected for study. Thus, MSU's conclusions do not necessarily apply to all species living in the study area or to changing future conditions. While the large amount of field data is a **strength** of the MSU studies and the objective of comparing risk assessment methodologies is admirable, the more narrowly defined purpose, lack of a comprehensive conceptual model, and statistical concerns in the resulting analyses **limit** the applicability of MSU's conclusions.

The receptor species and exposure pathways chosen for the conceptual model of a risk assessment are critical, and in this case vary significantly between the BERA and MSU studies. The terrestrial receptor species considered in the BERA were muskrats (a linkage between the aquatic and terrestrial food webs), earthworms, deer mice, robins, red fox, wood thrushes, yellow warblers, great horned owls, and red-tailed hawks. A number of terrestrial species were chosen by MSU for ecological studies (earthworms, other terrestrial invertebrates, herbivorous and omnivorous small mammals, shrews, bluebirds, house wrens, and great horned owls). However, the actual studies conducted by MSU emphasized those species that could be sampled easily in the field, primarily through trapping or artificial nesting structures, rather than their potential importance as identified in a comprehensive conceptual model. The choice of these species was sensible given logistical constraints related to obtaining sufficient sample sizes (i.e., abundant species) and accessibility (i.e., easily sampled species), and on account of MSU's main objective to compare lines of evidence derived from ecological sampling and food chain modeling. The inclusion of shrews and terrestrial invertebrates (other than earthworms) in MSU's studies was a significant addition to the BERA. Shrews are likely to experience elevated PCB exposure through their ingestion of earthworms, terrestrial invertebrates, and incidental soil. However, as discussed previously, MSU recently provided the Panel with data from shrew studies, but the limitations in those data, and the sparse associated earthworm data, make it unclear at this point how much those studies improved the understanding of the ecological risks. Nevertheless, the Panel does recommend that the shrew data be considered as a component in the multiple-lines-of-evidence approach, both as a pathway for exposure to higher trophic levels and as a representative of the receptor class of small mammals, and conclusions based on the results should be taken as far as the data and uncertainties allow in informing risk management decisions.

Although MSU's studies included to some degree all of the species listed in the BERA for food-web exposure modeling, hazard quotients were calculated only for bluebirds, house wrens, and great horned owls. These species do not necessarily represent the most highly exposed or the most sensitive species present in the riparian corridor. For example, one of the MSU papers acknowledges that the two passerine species "were not chosen as surrogate species or sensitive sentinels for other species but rather were studied to determine the potential for exposure of these species and to determine ecologically relevant reproductive parameters" (Neigh et al. 2007). By failing to characterize additional receptor species, the MSU studies did not focus on all of the important pathways of exposure that a comprehensive conceptual model would have identified. This is a major **limitation** to the conclusions that may be drawn from the MSU studies concerning overall ecological risks.

Likewise, MSU used relatively high avian TRVs, based largely but not exclusively on studies of less-sensitive passerines. While this approach may be appropriate for comparing to multiple lines of evidence for the specific species that were studied, it does not account for potential risk to potentially more sensitive species present in the floodplain ecosystems. Hence, MSU's risk conclusions apply only to the particular species that were studied and are not necessarily broadly protective of entire the avian community.

The soil—earthworm exposure pathway was a particularly important factor in driving a conclusion of potential risk for robins in the BERA. However, the MSU team collected few field data for robins. This was perhaps understandable if, in fact, robins fed rarely in the Trowbridge study area. At the May meeting and in Neigh et al. (2006a), MSU reported a dearth of earthworms in the Trowbridge study area. (Note: this observation is at odds with a) the earthworm sample sizes reported Blankenship et al. 2005 and b) the apparently common occurrence of earthworms reported by CDM for the BERA per DEQ memo.) Earthworms were not found in food bolus samples from bluebirds and house wrens at the Trowbridge site (Neigh et al. 2006a), so ecological studies and hazard quotients related to these two species, while important, provide no information about risks to earthworm-eating birds. Further consideration of the soil—earthworm exposure pathway is essential for evaluating risks to earthworm-eating wildlife such as robins, woodcock, snipe, and shrews.

The delineation of the time frame under consideration is also important to the issue of an adequate conceptual model —the charge asks for an evaluation of MSU's data and analyses as appropriate lines of evidence for future risk management decisions. At the May 13 Charge Meeting Dr. Giesy specifically stated that none of the MSU studies were designed to look at future scenarios, such as changes in habitat use with succession. The emphases of MSU's published studies are consistent with that statement, comparing multiple lines of evidence only for the species studied under present conditions. A time scale of many decades is more consistent with the ecological time scales considered by ecosystem managers. This is particularly to be expected if current plans by the State are implemented to completely remove the remnants of the Trowbridge Dam and other water control structures, resulting in altered hydrology in critical areas of concern. This would directly reduce the frequency and magnitude of episodic flooding events, which cause 1) the inundation of the formerly impounded areas with PCB-contaminated sediments, and 2) the erosion and exposure of contaminated soils. An important consequence of such a decreased frequency of inundation may be a significant increase in the density of earthworms in those areas. Thus, earthworm abundance and bioaccumulation of PCBs may change as soils develop over time in the formerly impounded areas, enhancing the magnitude of what was found in the BERA to be the critical pathway for risks. Failure to consider plausible changes in future conditions is a **major limitation** of the MSU studies.

4b. Data interpretation, including the choice and application of statistical methods.

Panel's Response -- Inconsistent and Unclear Design and Analysis Methods: The MSU Sampling and Analysis Plan (SAP) and Standard Operating Procedures (SOP) in Attachment #4 are in conflict with several methods reported in the published papers. One would assume that the methods reported in the published papers are correct; however differences should be explained and justified. For example, the SAP, page 68, calls for stratified random sampling with a random sample of size 2 from each of 3 strata for sediment, soil, and biota collection, or 6 locations total. The methods in the published papers indicate that 4 study sites were subjectively selected in the Trowbridge impoundment and 2 were selected in the Fort Custer reference site. Subjective selection of sites was reinforced in oral remarks made by Dr. Zwiernik during the Panel's tour of the Trowbridge impoundment site.

The Panel notes that the null and alternative hypotheses stated in the SOPs should essentially be switched with each other. Fortunately, this error does not appear to have found its way into the published papers or to have influenced the formulas for exploring sample sizes required to detect important effect sizes.

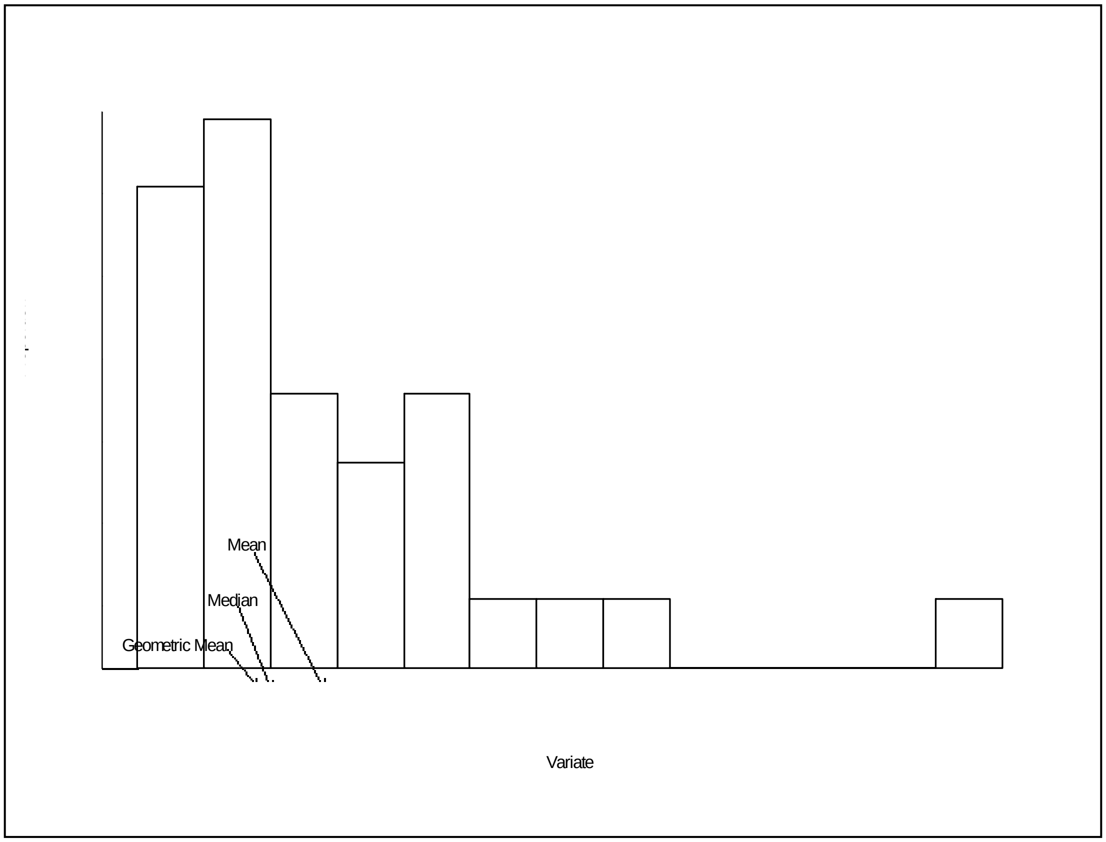
Panel's Response – Inconsistent Statistical Methods: Geometric means were used as a measure of central tendency in the distribution of concentrations of PCBs in HQs for great horned owl studies. Arithmetic means however, were used for computing HQs for blue birds and house wrens. We assume the upper 95% confidence limit used in computing HQs for the great horned owl is for a confidence interval on the geometric mean. First, methods should be consistent between the papers published by some of the same authors. Second, the geometric mean may not be appropriate for calculation of HQs, because for a sample from positively skewed PCB data, the geometric mean will be less than the median and the median is less than the mean, i.e., more than 50% of the sample data will be greater than the geometric mean. The geometric mean provides the smallest measure of central tendency (for positive skewed data) and therefore is least conservative and has the highest chance of a "no risk" decision (e.g., assuming we are talking about an exposure distribution or a distribution of metrics where large numbers are 'bad'). See the attached Figure 3.1, for the location of the geometric mean, median, and mean in an example sample of size 30 from a lognormal distribution.

Note that for the lognormal distribution, the "middle" value is estimated by the median. Therefore, the Panel suggests that the median or arithmetic mean of original data (not log-transformed data) be used for all analyses calling for measures of central tendency. The arithmetic mean will provide the most conservative estimate of central tendency of concentration of PCBs in food items because extremely large concentrations of PCBs are encountered, although with small probabilities, in lognormal-distributed data with their positively skewed values. The median or arithmetic mean should be used with this explicit understanding (e.g., they provide the middle, or the most conservative estimate of central tendency of PCB concentrations; also see Appendix A).

The choice of statistical estimate for central tendency should reflect a scientific consideration of the biological endpoint under consideration. For example, PCBs in a dietary item (e.g., worms) are probably represented by a lognormal-shaped distribution, where the shape of the distribution represents the expected probability of occurrence and variability of worm PCB concentrations over a defined geographical area in the Kalamazoo study area. The median (the 50^{th} percentile) of the measured PCB concentrations is generally an appropriate indicator of centrality for this biological metric. However, in the case of measured bird body burdens (representing bioaccumulation of PCBs over many feedings), the best statistic for central tendency may be the arithmetic mean. In particular, if the toxicological process associated with PCB diets over time is proportional to the total amount of PCB consumed, then the arithmetic mean is generally a better estimate of the effect of PCBs to the bird. In this case, even though the chance of encountering a high PCB concentration in the diet is relatively small, the proportional effect on the bird is equivalent to many feedings consisting of smaller concentrations. The arithmetic mean will therefore reflect the proportionality of the toxic effect. The Panel encourages both USEPA and MSU to provide a scientific and statistical defense for the selection of a statistical estimate of central tendency, with the narrative providing the reader with a rationale for the choice of statistic.

In addition, the Panel notes that the MSU papers did not present a laboratory-derived dose-response relationship for either total PCB or individual PCB congeners. A comparison of the measured PCBs concentrations in either dietary endpoints or body burdens to the shape and scale of the dose-response relationship may provide insights into the selection of an appropriate statistic representing central tendency. For example, a finding that the majority of the measured data falls in the lower (or upper) portion of the dose-response curve will provide insights into the influence that any one measurement has on the interpretation of the central-tendency statistic. The Panel encourages MSU to provide such graphics and analyses in their defense of the central-tendency statistic.

Figure 3.1 Geometric mean, median, and mean for a sample of size 30 from a positively skewed lognormal distribution



Panel's Response – Pseudo-Replication Issues (Also see Appendix A): It is not possible to make design-based statistical inferences to the entire Trowbridge impoundment and the entire Fort Custer reference site as implicitly implied, because at best, there appears to be a subjectively selected sample of size one from each of four strata in the Trowbridge impoundment and a subjectively selected sample of size one from each of two strata in Fort Custer. Statistical inferences to Trowbridge impoundment and Fort Custer are based on pseudo-replication (e.g., resulting in non-representative data) within the relatively small sampling grids and clusters of nest boxes. Statistical results are at best limited to the combined areas of the sampling grids and clusters of next boxes; however, the MSU papers did not appear to check that 'replications' within sampling grids or nest boxes within clusters were sufficiently far apart in space or time to ensure that they were statistically 'independent.' Potential spatial and serial correlation within the sampling grids and within clusters of nest boxes would tend to decrease the effective number of replications and hence increase the standard errors of estimates, increase the width of confidence intervals, and increase the p-values of statistical tests, thereby decreasing confidence in results and conclusions by some unknown amount.

Panel's Response – The Use of Finite Population Corrections Factors for Nest Productivity Studies: The Panel was asked by KRSG to consider whether or not it is appropriate to use a Finite Population Correction Factor if a large proportion of the population of nesting bluebirds, house wrens, and great horned owls living in the study areas were sampled within those boundaries. In response, the Panel assumes that KRSG is referring to population correction factors that applied to variance estimates when the sampled data represent a large proportion of a known population (i.e., a large proportion of bluebird nests in the geographic area of interest). Note that all conclusions in the MSU papers reviewed by the Panel refer to the entire Fort Custer area and the entire former Trowbridge impoundment, not the specific areas encompassing the sampled nest boxes. The areas encompassing the nest boxes were not selected by a probabilistic sampling procedure; therefore, they do not provide statistical inferences to the entire study area. Conclusions made in the MSU papers reviewed by the Panel were written in the spirit that they apply to some 'sample' of a large potential population of great horned owls, bluebirds, and house wrens that could have nested in the study areas. Therefore, even if MSU did sample a large percentage of actual nests in a given area, the use of a population correction factor is inappropriate because the data are interpreted as being representative of a larger geographic region.

If KRSG wants the inferences to be limited to the finite nesting populations in the sampled region, then all analyses must be stated in terms of point estimates with a measure of 'error of estimation.' The Panel again emphasizes that all tests of hypotheses, confidence intervals, and related inferences using a finite population correction factor be removed from the papers. In fact, the Panel has emphasized that the use of hypothesis testing as a means of decision-making with any biological endpoint is ill-advised. The finite population correction factor does not apply to these MSU papers as written.

Panel's Response – Does Use of Nest Boxes Negate Habitat Differences Between the Former Trowbridge Impoundment and Fort Custer and Make Pseudo-replication a Moot Issue; One source of pseudo-replication is that study units (e.g., nest boxes) and sampling grids (for collection of soil, etc.) were not selected using a probabilistic procedure (e.g., simple or systematic sampling of points for location of nest boxes) from the *entire study areas*. This form of pseudo-replication implies that conclusions are not necessarily valid for the entire study areas. If the conclusions were clearly limited to the areas containing the nest boxes, then this source of pseudo-replication would be eliminated. However, other sources of pseudo-replication still exist (i.e., nest boxes were not located by a probabilistic procedure within the 'smaller area' containing the next boxes). Therefore, the use of nest boxes does not negate issues of pseudo-replication that are described elsewhere in this Report.

Panel's Response - Application of Spatial Statistics to Test for Independence of Bird Nest Box Productivity Results and Worm Samples: The Panel was asked by KRSG to consider whether or not spatial statistics could be used to establish the independence of replicates within sampling grids or nest boxes as a means of concluding that the nest box clusters were independent in space or time (thereby negating the issue of pseudo-replication on data interpretation). If spatial statistics are applied to test the independence of the replicate data, and the analysis concludes the data are indeed independent, will pseudo-replication remain an issue? Can such tests of independence also be conducted with the worm samples to see if the clusters of samples in space and time can be treated independently? Analyses could be conducted in the spirit of "geostatistics", i.e., now better known as "spatial statistics," if there are sufficient data to estimate the semi-variograms required by all geostatistical procedures and associated tests of spatial independence. However, as noted above, inferences must be clearly limited to some undefined small areas containing the nest boxes (and foraging activity) and within the small sampling grids for soil, etc., **not** to the entire impoundments and reference area. Geostatistical inferences to a larger part of the impoundment or reference area might be possible for the nest box data if the sample size is considered to be the number of nests. However, there are clearly **no** data to conduct a geostatistics analysis at the larger scale, and based on the already described issues with pseudo-replication and survey design, extrapolation from the nest box data to larger scales is ill-advised. Similarly, samples within the worm/soil sampling grids are quite limited.

Panel's Response – MSU Findings of No Effect: All of the studies have an over-dependence on use of statistical tests of null hypotheses. Acceptance of a null hypothesis of 'no difference' between sample sites or times is inappropriately used to conclude that the parameters from the two sites or times are in fact 'equal' or that data sets can be pooled. It is widely known that acceptance of a null hypothesis does not necessarily imply that there are no important biological difference between the parameters under consideration. The summary presented to the Panel by Drs. Giesy and Zwiernik (both oral and written) at the meeting at Brook Lodge made excessive use of inappropriate acceptance of null hypotheses in attempting to support their conclusions. Furthermore, "no effect" is dependent upon sample size and variability in the data; thus, one might argue that a more robust design with larger sample size (particularly for GHO and bluebirds) could return different results. See Appendix A for further comments on recommended statistical analyses.

4c. Approach for addressing natural variability

Panel's Response: There are two aspects of natural variability that are relevant here: within-site variability and between-site variability. Within-site variability refers to variability of exposures and effects within a specific impoundment, e.g., Trowbridge. Between-site variability refers to variability of exposures and effects between impoundments. The MSU studies did not directly evaluate either type of variability. Information on PCB concentrations in food items, eggs, and avian tissues are summarized in the MSU papers as means and standard deviations. All further analyses (e.g., dose modeling and HQ calculations) are then based only on the means. It would be relatively straightforward to use these data to quantify the variability of exposures within either of the two sites studied, still recognizing that significant issues exist with small sample sizes and pseudo-replication. For example, the means and standard deviations of PCB concentrations in prey items could, through Monte Carlo analysis of an assumed dose model, be used to compute the distribution of doses within the bluebird and house wren populations inhabiting the Trowbridge and Fort Custer sites. From these assumed distributions or by re-sampling the sample data with replacement, the fraction of birds potentially exposed above any given level (e.g., all of the selected TRV dose) could be computed (GE 2005). Similarly, estimates of the means and variances of PCB concentrations and eggs could be used to conjecture distributions of resulting egg concentrations. A complete probabilistic assessment could be based on these assumed distributions by using a range of TRVs rather than a single number, and using Monte Carlo techniques to compare the range of exposure values with the entire range of TRVs. This would essentially obviate the need for deciding which single TRV to use, and instead will compute the HQ probabilities associated with the assumed distributions of exposure dose and effect level. Acknowledging the problems with small sample sizes and pseudo-replication, these model-based analyses would be informative.

Since only two sites were studied by MSU, relatively little information is available about between-site variability. As noted elsewhere in this review, information on between-site differences in diet compositions could be obtained by comparing diets of birds nesting at the Trowbridge and Fort Custer sites. This information could provide a bound on between-site variability of diets at other sites, because the Fort Custer site appears to be ecologically quite different from any of the contaminated impoundments.

4d. Identification and characterization of uncertainties.

Panel's Response: Neither the BERA nor MSU studies conducted a formal uncertainty analysis in their respective approaches to establishing ecological risks. Given the differences in risk characterization among the studies, the Panel strongly recommends that the BERA and MSU data have formal uncertainty analyses performed. MSU, for example, could generate exposure and effects distributions using probabilistic techniques (rather than the simple hazard quotients) and model-based analyses, while acknowledging the limitations of pseudo-replication and small sample sizes. Similarly, simple model-based sensitivity analyses of hazard quotients formed with differing estimates of exposure and effect could be implemented and graphically presented. (See Appendix A for discussion of approaches to conducting uncertainty analyses). Other sources of uncertainty have been discussed by the Panel elsewhere in this Report, with the overall conclusion by the Panel that there has been insufficient identification and characterization of uncertainties, and insufficient attention to the implications of the uncertainties to the weighing of results and conclusions that could be drawn from the MSU studies.

4e. Adequacy of the data to support inferences on population-level effects.

Panel's Response: The MSU productivity studies did not measure all the parameters necessary to develop a population model or to make predictions about long-term changes in per area density of the various bird species (either as individuals or as breeding pairs). This apparently was not the intent of the studies, which were designed to specifically address questions about PCB-related changes in productivity. It frequently is implied that a statistically significant reduction in productivity will lead to population declines, but this is truly an untested hypothesis for each species. Compensatory or density-dependent changes in juvenile or adult survival may off-set reduced productivity, resulting in similar densities over time. Of course, the age structure of the population could change to an older aged population if productivity, but not survival, were affected. The **strengths** of the MSU studies are the site-specific productivity metrics that were developed (for some of the species). These potentially could be used in a simple matrix population model-based analysis, using stage-specific survival data from the literature (preferably from studies in the same geographic area). The **limitations** of the MSU data are the small sample sizes and study design difficulties. Those data (such as for the bluebirds) that are sufficiently robust to provide a reliable estimate of fledging success per nest would be very useful in simple population models; this will allow a sensitivity analysis to be conducted to confirm which life stages(s) are most affected and how this will impact the population over the long term.

Panel's Response: MSU's assertion that one "bad year" and two "good years" is evidence that contaminants are not affecting productivity is not supportable. There is evidence from other studies that interactions of other environmental stressors (e.g., climate, parasites) with contaminants can result in reduced productivity, while the same contaminants in otherwise "good years" will not (Nagy, L, N. Schumaker, and A. Fairbrother, unpublished data on western bluebirds in the Willamette Valley, OR). Furthermore, apportioning causes of nest failure is difficult, although methods now are available to provide a more quantitative estimate (e.g., Etterson et al. 2007). A good example of a true demographic (population) study of a passerine is a study of the western bluebird by Keyser et al. (2004).

Call counts provided indices of population density for great horned owls, but these indices may have been influenced by the use of artificial nest platforms. This habitat manipulation significantly reduces the strength of conclusions (e.g., no effects of PCBs) that can be drawn from these population indices.

The study of metapopulation dynamics moves beyond population density to examination of recruitment rates and landscape-level spatial patterns (e.g., source/sink population dynamics). In ecotoxicology, a key question related to metapopulation dynamics is whether young raised in that area survive and reproduce, either in that area or elsewhere. A previous study by the MSU lab showed significantly low recruitment rates of young Caspian terns raised in ecosystems contaminated with PCBs and other organochlorine contaminants (Mora et al. 1993). Likewise, the MSU lab showed that bald eagles nesting on contaminated Great Lakes shorelines had reduced productivity of young, which contributes to a lower recruitment rate from those areas (Best et al. 1994). In both situations, the numbers of breeding adults in the contaminated areas were supported by the immigration of birds raised in less contaminated areas. The Kalamazoo River studies did not assess recruitment rates and migration patterns, and hence potential effects of PCBs on metapopulation dynamics remain unknown.

Panel's Response – Great Horned Owl Population Estimates: The nest productivity data address organism-level effects but do not support inferences concerning population-level effects of PCB exposures. The great horned owl call-count surveys conducted by MSU potentially address population-level effects; however, the Panel believes that MSU's conclusions concerning those effects are not supported by the data.

Call-count surveys of great horned owls were conducted for two purposes: 1) locating owl nests for studies of tissue residues, reproductive productivity, and dietary composition; and 2) as indirect measures of relative population density. Two of the MSU papers use the call counts as lines of evidence of a higher owl population density at Trowbridge compared to the Fort Custer reference, supporting a conclusion of lack of population-level effects at Trowbridge (Strause et al. 2007a, Zwiernik et al. 2007). While this conclusion based on call-count indices is consistent with the higher number of occupied nests at Trowbridge, inconsistencies and incomplete explanations in MSU's descriptions of the call-count methods raise significant questions about specific (i.e., numerical) comparisons with previously published and potential future studies. The SOP (272) and published papers appear to be inconsistent with respect to cited references and methodological description. The SOP cites Frank (1997) as the source of the methods. Strause et al. (2007a) cites Brenner and Karowski (1985) and Zwiernik et al. (2007). Zwiernik et al. (2007) cites both Frank (1997), Brenner and Karwoski (1985) and Rhoner and Doyle (1992). These citation questions might seem irrelevant but for the different descriptions of the methods in the SOP and papers. The SOP describes evening surveys with call locations at 0.5 km intervals and pre-broadcast, broadcast, and post-broadcast periods at each location. Strause et al. (2007a) describes an "active" method (the term *active* not defined) with morning and evening surveys at call locations at 0.5 km intervals. Zwiernik et al. (2007) describe both an active survey method with hoot broadcasts and a passive or silent survey during sensitive life stage events (these events not defined but Rhoner and Doyle [1992] was cited). Given the variability in the above descriptions, it is difficult to determine how data for various observation methods and periods were used to calculate the response rates in the categories of "total," "foraging," and "paired" used in both MSU papers. (Note that MDEQ's concerns over potential biases in the call-count methods are difficult to evaluate because the description of MSU's methods are so unclear.)

Overall, these inconsistencies raise the questions of exactly what was done and whether the protocols were compatible with well-accepted methods and previously published data for great horned owls. A specific example of such a (potential) comparison is the statement in Zwiernik et al. (2007) that "measures of site-use (abundance) indicated the target area populations at

Trowbridge were near the carrying capacity for undisturbed GHO habitats (Houston et al. 1998)." This statement is unclear whether the "site-use (abundance)" comparison is being made based on 1) call-count indices of relative abundance (responses per survey) generated using identical field protocols in both studies, or 2) estimates of breeding population density (number of breeding pairs per unit area) using call counts and many other observational methods to identify all breeding pairs in an area. In either case, identical or compatible methods would have to be used in both studies to allow specific numerical comparisons to be made.

Beyond these issues of call-count protocol, the usefulness of the relative abundance indices is severely limited by the insufficiency of the Fort Custer as a reference site. Call counts using the same methods should be applied to other (and replicated) references sites in southwestern Michigan or the upper Midwest to determine the magnitude of and variability in these indices in healthy populations.

3.2.5 Question 5: What are the relative strengths, limitations, and associated uncertainties that should be considered when extrapolating from the results of MSU studies conducted in the former Trowbridge Impoundment to the other formerly impounded areas of the Kalamazoo River?

Panel's Response: Given the limited amount of habitat- and spatial-related information provided to the Panel, there is considerable uncertainty in extrapolating results from the Trowbridge impoundment to other formerly impounded areas. The MSU papers and reports do not adequately describe the relationship between habitat and exposure/effects data. A spatial re-analysis of MSU's data might be insightful, but in some cases the samples sizes for particular types of samples are quite small or data were composited across sampling locations, severely limiting their use in habitat-specific analyses and extrapolations. While the types of habitat and plant communities found in the former impoundments appear to be generally similar, they do appear to differ in their relative distribution, and perhaps in other important characteristics including patch size and connectedness, which may affect the conclusions that could be drawn from the studies. A more definitive description of future land management goals by the Trustees would also help clarify these extrapolation questions.

Supplemental Issues to Consider:

5a. Numeric and spatial distributions of PCBs in floodplains of former impoundments

Panel's Response: The numeric and spatial distribution of PCBs in soil and biota is poorly described in MSU's papers and reports. Fig. 1-2 in MSU's May workshop report (Giesy and Zwiernik 2008) show soil/biota sampling and nest box/platform locations, but no subsequent information is provided linking sampling locations with numeric PCB data. Given the low sample sizes for some sample types and issues with pseudo-replication, conclusions regarding spatial (and temporal) variability are likely to be significantly limited. Note that Chapter 7 of the

Stage 1 NRDA assessment (Stratus 2005) does a more complete job of characterizing risk on a spatial basis and may provide an example of what could be done with MSU's data.

5b. Habitat characteristics in floodplains of formerly impounded areas

Panel's Response: Factors to be considered when making extrapolations of Trowbridge data to other impoundments include variations in habitat type, food-web structure, soil/sediment PCBs, and likely utilization by various passerine species. At a minimum, impoundment-site-specific conceptual models will be needed to identify the key uncertainties relevant to each impoundment. To account for future changes in diet composition, these conceptual models should also include changes in habitat characteristics related to ecological succession.

The obvious ecological differences between Fort Custer and Trowbridge could be used to characterize between-site differences in passerine diets. The MSU publications combine diet data for each species over all sites. Although this approach is useful for comparing the diets of different species over a range of habitats, it obscures within-species differences in diets that may occur caused by differences in habitat quality or prey availability at different sites. Within-species comparisons of diet compositions of birds nesting at Trowbridge to diet compositions of birds nesting at Fort Custer would permit at least a qualitative evaluation of the influence of site characteristics on passerine diets. Supporting reference: Neigh et al. (2006a), which contains diet composition data for tree swallow, house wren, and eastern bluebird, for Trowbridge and Fort Custer sites combined.

Note: The Panel's ability to respond to Charge 5b has been limited by the availability of site specific habitat-related information for the former impoundments. The MSU papers and reports contain little information on the spatial distribution of sampling, including the relationships between samples and particular habitat types. Habitat information from the Trustees has also been minimal. Some of the most detailed information on habitat is found in Section 3 and Figures A-1 through A-5 in ARCADIS, 2008 report provided to the Panel at the May 13 Charge Meeting. The Panel was told that this information had not been vetted by the Trustees. Examination of the habitat maps shows generally similar habitat types present in the former impoundments. Comparisons of the degree of habitat fragmentation are difficult given the different scales used for some habitat maps.

5c. Likely utilization of floodplains in formerly impounded area by the receptor species evaluated in MSU studies

Panel's Response: Since the mix of habitats at Trowbridge seems to be generally similar to the mix at the other impoundments, one would expect to find similar receptor species present. The particular receptor species chosen for study by MSU are relatively common species for this region of Michigan, and hence would be expected to be present if appropriate habitat is available. Great horned owls might be one exception--the smaller impoundments would likely be big enough for only a few (or even a partial) owl territory. Populations of cavity nesting passerines such as bluebirds and house wrens would be limited by the abundance of natural cavities (e.g., in dead trees) in the absence of nest boxes. Relatively simple, qualitative wildlife survey's (e.g., visual bird observations or call counts, limited small mammal trapping) could be used to clarify uncertainties in the animal community composition at the other impoundments.

3.2.6 Question 6: Please comment on the applicability of the information presented in the MSU studies for informing risk management decisions.

Panel's Response: The applicability of the information presented in the MSU studies for informing risk-management decisions depends upon the following four considerations:

- Data quality: conformance to USEPA standards for sample collection/handling, analytical chemistry, database management, etc.
- Study design: species/site selection, selection of metrics, sample size, as function of study objectives.
- Relative value of empirical studies performed by MSU (soil/biota concentrations; nest productivity; analysis of PCBs in nestling diets) vs. literature-based analyses (TRVs)
- Interpretation of results: Conclusions supported by data?

With regard to data quality, it appears to the Panel that MSU followed USEPA's recommended procedures. With regard to study design, as noted elsewhere in this review, MSU's approach was significantly narrower than the approach taken in the BERA; moreover, one of the key receptors (robin) evaluated in the BERA was not addressed by MSU. In addition, the nest productivity studies were compromised by small sample size, pseudo-replication, and lack of comparability between the Trowbridge site and the Fort Custer reference site. The other empirical studies performed by MSU, specifically the measurements of PCB concentrations in soil and biota, the direct measurements of PCB doses to nestlings, and the quantification of species-specific dietary preferences, did not suffer from these flaws. These empirical studies provide new data that can be used to support refined area-specific risk assessments and other studies performed to support the risk management process at this site. MSU's approach to TRV selection does not appear to be superior to the approach used in the BERA and provides no new information for risk management. Because conclusions concerning risks presented in MSU's published papers are heavily dependent on values chosen for TRVs and the justification for selecting specific TRVs is inadequately described, the Panel believes that MSU's risk conclusions are not supportable. However, the risk assessment approach used in the BERA could be modified to accommodate MSU's site-specific exposure data, thereby significantly enhancing the quality of risk information available to risk managers.

In this regard, the Panel notes that neither the BERA nor MSU implemented any sort of formal uncertainty analysis in their respective approaches to establishing risk. Again, given the differences in risk characterization among the studies, the Panel strongly believes that formal uncertainty analyses should be conducted to support any future use of the MSU data (acknowledging limitations of small sample sizes and study design) and for any other data used in risk management. For example, model-based exposure and effects distributions could be generated using probabilistic techniques (rather than the simple hazard quotients). In addition, simple model-based sensitivity analyses of hazard ratios formed with differing assumed values of exposure and effect could be implemented and graphically presented. Uncertainty bounds on the resulting clean-up values could be generated, subject to assumptions made and limitations of the data. A comprehensive listing of mathematical approaches for conducting these model-based analyses is not presented here, but can be found in Warren-Hicks and Moore (1998) and Warren-Hicks (1999).

4.0 Panel's General Comments

The studies conducted by the MSU team provide additional valuable information to inform the ecological risk assessment and risk management decisions on the Kalamazoo River. The **strength** of their work lies in the site-specific data collected that can be used to verify the dietary exposure models used in the BERA and provide additional lines of evidence. The additional lines of evidence include egg concentrations of PCBs and productivity measures of the study species. The **limitations** of the studies include: inadequate statistical design; insufficiency of many of the data; absence of a comprehensive conceptual model relating exposures and effects on endpoints of concern; the lack of detailed information in the publications (i.e., the lack of a study report that could contain much more detail than allowed in a literature paper); inadequate documentation and justification of the selected TRVs; inadequate identification and quantification of sources of uncertainty; and the over-interpretation of the results provided in MSU's 2008 summary document.

While acknowledging the small sample sizes and issues with study design, the best use of the MSU study results, would be to:

- Use the site-specific tissue data in the dietary exposure models in the BERA – the **strength** of this approach is to provide site-specific BSAFs and BMFs and measured concentrations in biota, rather basing the food-web model entirely on literature-based estimates. This will incorporate soil-specific effects (e.g., soil carbon), congener-specific differences in accumulation rates, and species-specific information related to the site (particularly for raptors, where literature-based data are very sparse). The **limitation** is that PCB concentrations in earthworms were inadequately measured, so the robin exposure pathway cannot be verified. Therefore, the estimate in the BERA will need to stand as the best assessment of risk to robins, although it may be modified/strengthened by site-specific adjustments of BSAFs used to estimate earthworm concentrations.

- Use the "bolus" data from the avian nesting study to further verify dietary exposure estimates **–** the **strength** of this approach is that the food bolus represents precisely what the nestlings are eating. By comparing the concentrations in this bolus to the estimated concentrations from the dietary exposure model, the model can be further refined to accurately reflect the diets and exposures (BSAFs/BMS) of the studied species. This may provide some additional realism for extrapolating to the non-measured species, such as the robin. The **limitations** of this approach is that only the house wren is truly feeding on only terrestrial foods, while the eastern bluebird, the tree swallow, and the great horned owl access some (or most) of their diets from the aquatic food chain. Thus, relating diet to soil contamination alone will be difficult.

- Use the great horned owl dietary assessment (Strausse et al. 2008) as the input to the exposure assessment for raptors at KRSS. A **strength** of this study is the direct measurements of PCBs in some GHO prey items that rarely are analyzed, and a reasonable comparison between KRSS and the reference site. However, it should be pointed out that PCBs were not measured by the MSU studies in rabbits and large squirrels, which represent 50 to 75% of GHO diet on a mass basis. Further strengths are the presentation of data on both a mass- and a concentration-basis, plus inclusion of both means and 95% UCLs of the means. However, until agreement is reached on appropriate TRVs, the hazard assessment presented in the paper should not be used.

- Studies of productivity of the bluebirds, wrens, and great horned owls provide useful, qualitative evidence of reproductive performance of on-site species. The **strength** of these studies is that they are directly measuring one of the assessment endpoints ("do PCBs affect reproduction of birds?"). Field measurements are usually preferable to laboratory-based studies as they include much more realism, including the fact that contaminant-induced changes are not always additive to other stressors that can reduce productivity (weather, predators, etc.). Of course, this is a limitation in field studies as well, as it requires a large sample size to be able to apportion causality to observed effect and to statistically show differences among local populations. The **limitations** of the study are the small samples sizes, issues with pseudo-replication and other aspects of study design, reliance on aquatic organisms for a portion of the diet of the bluebirds and owls, lack of accounting for observational artifacts (such as time of nest initiation or failure) with the great horned owl study, and the confounding effects of habitat differences among the KRSS sites and the reference area (Fort Custer). Further complicating the interpretation is the bluebird boxes have been on-site for years at Fort Custer but were newly erected at Trowbridge; box use is known to be significantly affected by familiarity of the birds with the placement of the boxes. Nevertheless, these studies can be used in a qualitative manner to relate site productivity with generally expected reproductive success of the species within the region.

- The MSU data can also be used to build models linking measurements of dietary intake to body burden. In addition, the MSU data can be used to evaluate relationships between measurement endpoints over space. Each of these approaches should provide insights currently not found in the BERA.

The MSU studies should not be used to reach risk conclusions on their own. There is too much uncertainty underlying the data interpretation, lack of robustness in the study design, and insufficient documentation (and lack of agreement) of TRV derivation. Some of the papers are repetitive (e.g., Strausse et al. [2008] and Zwiernik et al. [2007] both describe risk to great horned owls using essentially the same data). One study (Strausse et al. 2007) on relationship of PCB concentrations between nestling blood plasma and eggs in great horned owls and bald eagles is interesting and provides good information for future monitoring studies, but is not particularly relevant to the current risk assessment at KRSS. Otherwise, the papers each contribute some data and information that can be used in an assessment of risk if integrated with the data and approaches used in the BERA.

In summary, MSU studies provide useful data for quantitative exposure estimates and qualitative weight of evidence for estimating effects. They can contribute information to provide a more comprehensive assessment of risk than currently provided in the BERA. However, the limitations of these studies indicate that they should not be used as stand-alone documentation, and the conclusions presented based on these data are not supportable.

5.0 Panel's Recommendations – Looking Forward

5.1 Major Recommendations

1. The Panel recommends that the **MSU conclusions** not be used to reach risk conclusions on their own. There is too much uncertainty underlying the data interpretation, lack of robustness in the study design, and insufficient documentation (and lack of agreement) of TRV derivations. However, the **MSU study data** when combined with data from the BERA can be useful to inform the ecological risk assessment and risk management decisions associated with the KRSS.

2. The Panel recommends that, given the complexity of the datasets developed by USEPA and MSU, a multi-party technical working group consisting of scientists (biologists, risk assessors, statisticians etc.) representing USEPA, MDEQ, and the KRSG should be established to oversee the conceptual model development, cross-comparisons, uncertainty analyses, and dataset synthesis activities needed to integrate the MSU data with the dataset used for the BERA. Further, the Panel recommends that any future ERA activities (e.g., area-specific risk assessments) performed to support remedial action decisions at the Kalamazoo River site should be developed cooperatively, using a Data Quality Objectives (DQO) process based on USEPA guidance and other applicable documents (USEPA 1994, Barnthouse and Suter 1996). Following the DQO approach, the technical working group discussed above would develop the comprehensive conceptual model and identify key receptors of concern for the ecological risk assessments. Once these steps have been taken, the group would review any remedial decisions that have already been made (e.g., regarding cleanup goals or remedial technologies for sediment) and identify information needed to support further remedial decision-making. The group would then evaluate the integrated data set to determine whether it is adequate to support risk management decisions or if key data gaps still remain. The above steps, in essence, complete the Problem Formulation phase of an ecological risk assessment (ERA) (USEPA 1992, 1997). If the existing data are found to be adequate to support remedial decision-making, then the group would complete the remaining steps of the ERA. If significant data gaps are identified, then the group would continue the DQO process by identifying the specific decisions for which the data are needed, stating how additional data would support those decisions, and developing sampling plans to obtain these data.

3. The Panel recommends that the parties involved collaboratively develop a comprehensive conceptual ecological model (CEM) that captures the exposure pathways and species or other ecosystem attributes of concern. The USEPA ecological risk assessment framework and associated guidance (USEPA 1992, 1998), as well as the guidance for conducting ecological risk assessments for Superfund sites (USEPA 1997), describe the Problem Formulation component as a critical step that defines the problem at-hand. One key

aspect of Problem Formulation is the development of a comprehensive conceptual model characterizing the stressors (here the PCBs) of concern, the ecological attributes (endpoints) at-risk, and the stress-response relationships between the two. The purpose of such a CEM is to capture the scientific understanding of the ecosystem and the stressor(s) impinging on it. A well-designed CEM also identifies the important pathways contributing to the risk and the uncertainties associated with each pathway. The CEM can be used effectively to communicate among scientists and decision-makers, as well as to stakeholders and the public, the issues at hand, how the existing or planned studies and data map onto the risk assessment, and how the system may change over time in ways that may affect the risk assessment. The Panel has noted many instances where a comprehensive CEM would be useful to characterizing and understanding the risks and the potential efficacy of management solutions. Even though Problem Formulation and development of a CEM is ideally accomplished before the risk assessment in undertaken, that does not diminish in any way the utility of developing a comprehensive CEM at this point in the risk assessment and risk management process. To the contrary, the Panel believes this is a critical step that needs to be accomplished before decisions can appropriately be made based on the best available science concerning the Kalamazoo restoration.

The Panel recommends that in developing the CEM, the parties involved focus primarily on capturing the scientific understanding of how this ecosystem works and how it is structured, and *not* begin with a focus on the data available or studies that have been done. In developing the trophic structure aspect of the CEM, the focus should be on feeding guilds and functional components, rather than initially on species. Once the trophic structure is articulated, then it will be a straightforward process to map particular species onto that conceptualization, including properly placing species that overlap more than one trophic level or feeding pathway. By developing the comprehensive CEM in that manner, the utility of particular species to understanding the important components of the risk can be better understood, and the potential for substituting species or extrapolating across representative species can be recognized.

One possible avenue to pursue in developing the CEM is to explicitly couple the aquatic and terrestrial systems into an integrated CEM, i.e., capturing those attributes in the real-world that link the two habitats and provide routes of exposures to species of concern. The present artificial separation of the aquatic and terrestrial systems has presented issues to the risk management process, such as partial pathways to one species from both aquatic and terrestrial trophic linkages that have not been adequately resolved, as the Panel discussed elsewhere. If a comprehensive CEM is to be developed, explicitly recognizing the aquatic-terrestrial linkages is essential to fully understanding the risks and how they may change for species in one habitat as restoration in the other habitat proceeds.

The Panel also recommends that in developing the CEM, outside expertise be recruited into the process. This can bring in fresh perspectives on the nature of the risks, add important expertise to the mix already available, and provide a neutral forum for exploring the scientific issues and uncertainties surrounding scientific questions or differences in interpretations that currently exist. The incorporation of neutral, outside expertise, if successful, might also prove appropriate for other aspects of the risk assessment process beyond developing the CEM, such as in interpretation of data, characterizing uncertainty, or developing potential scenarios of future conditions.

Continuing the point about future conditions, one important contribution that a comprehensive CEM can make is an improved understanding of how alternative ecosystem conditions that may occur in the future could affect the risk and, therefore, the recovery process. While the base CEM could characterize the ecosystem as it presently is, alternative CEMs should also be developed based on a different hydrological regimes in the future (such as following removal of remnant dam structures) and the subsequent successional changes that would be expected to ensue. Thus, CEMs would capture the future conditions scenarios discussed elsewhere in the Panel's report and would provide a systematic basis for risk management decisions that are appropriate not only for the present, but also for plausible future conditions

4. The Panel recommends that the risk assessment process include a set of scenario-consequence analyses, in which a series of plausible future conditions are incorporated into the CEM (discussed above) and assessments are done of the resulting risks. The Panel recommends that this be accomplished in a nested approach: First, do the risk assessments based on the current hydrological regime and develop a few scenarios of the trophic structure and associated species of concern that exist now and at selected points-in-time as succession proceeds. Second, examine the risks for an alternate hydrological regime, such as based on removal of remnant dam structures and associated changes in water levels, frequency of inundation, etc.; the initial ecological structure under the altered hydrological regime should be assessed, followed by risk assessments on a selected set of subsequent trophic structures as succession proceeds. Next do the same, but based on a third hydrological regime, such as a wetter future climate, with associated succession stages. Other alternate hydrological regimes should be systematically examined as appropriate to capture the full range of plausible ecosystems that may occur over a reasonable period of time into the future, such as 100 yr.

5. The Panel recommends using a systematic approach incorporating all data from both MSU and BERA. Suggested activities include the following: (1) evaluation of the ability to pool the MSU and BERA data into a unified data set; (2) evaluation of the methods used to quantify the magnitude of PCB concentrations generated by each study (e.g., are the analytical chemistry results consistent among studies?); (3) calculation of uncertainty in the BERA results based on the extended data generated by MSU; (4) re-evaluation of

the MSU statistical results using alternative statistical estimates of central tendency of measures of PCB concentrations; (5) comparison of BERA and MSU results using formal uncertainty analysis methods; (6) evaluation of the effect of temporal variability on the MSU and BERA findings; (7) comparison of the BERA findings to model-based findings that could be generated using the MSU data (see Appendix A); (8) evaluation of the findings in light of the conceptual model employed by each study; and (9) evaluation of the BERA findings in light of the quantifiable relationships between PCB source concentrations and PCB egg and body burden concentrations that may be obtainable with the MSU data (note: these analyses were not implemented by MSU, but the MSU data set suggests that such analyses may be possible).

6. The Panel recommends that a cross-comparison between the MSU and BERA studies be made using exposure data from one in the model from the other. For example, it would seem to be a simple exercise to take the PCB concentration data in soils and lower-trophic level samples and run them through the BERA model to see what exposures to higher-trophic-levels would ensue, while acknowledging the issues with small sample sizes and study design. Similarly, the BERA data could be run through the MSU exposure model. The Panel has already performed a preliminary cross-comparison, as shown in Table 3.6 in this report. It is to be expected that the results will differ between the two studies because of differences in scopes and study designs. Nevertheless, such a cross-comparison could illustrate the magnitude of the differences resulting from the two approaches and the causes of the differences. If considered along with an improved understanding of uncertainties associated with each approach cross-comparison will inform the risk managers of the validity and defensibility of each set of analyses, and enhance their ability to appropriately weigh differences in results. This exercise would seem to be essential before any reasonable understanding of the multiple-lines-of-evidence approach could be reached from the two disparate datasets and results. Moreover, the MSU data, with measurements on the endpoints of interest, should not be limited to analysis using only the USEPA approaches. See Appendix A for other approaches, as well as additional quality assurance issues, that could be applied to the MSU data sets. Note, the recommendation for cross-comparison of MSU and BERA information is meant to be used in an exploratory assessment, not to develop final risk numbers. Additionally, this cross-comparison would not be appropriate for the productivity data developed by MSU.

7. The Panel recommends that, rather than focus on estimating a single risk number, the ecological risk assessments would be strengthened by presentation of a distribution of risk levels tied to the uncertainties in the underlying data and/or model structure (e.g., relative importance of different dietary pathways). Consequently, the Panel recommends that the exposure models and data in the MSU study and the BERA be subject to a formal uncertainty analysis. Included in this should be an extensive sensitivity analysis of the models to explore the plausible range of risks in the system. For example, one set of the

variables in an exposure model is the particular diet of an endpoint species. The frequency distribution of dietary sources could readily be varied across a large number of scenarios, allowing calculation of how sensitive the resulting assimilated dose is to the dietary composition. Similarly, use of different specific bioaccumulation factors within the range of plausible values for each could be explored in a set of Monte Carlo simulations. Other model structural and parameter sensitivity analyses would enhance the understanding of the ecological risks in this system and could suggest specific additional research needed to reduce uncertainties.

8. To address concerns about the approach used in the MSU study to develop toxicity reference values for the PCBs of concern, the Panel recommends that consideration be given to applying an approach that uses the full set of available, high-quality toxicity studies rather than a single study to derive a TRV. Such an approach could be modeled after the EcoSSL methodology USEPA developed for PAHs, the methodology used by USEPA to set water quality criteria (Stephan et al. 1985), or the methodology used by USEPA (2004) to develop the TRVs used in the Housatonic River BERA, among other examples. The Panel recognizes, however, that using a single toxicity study to derive a TRV is a part of USEPA-approved methodology, under certain circumstances, and therefore could be used in the ecological risk assessments here. Regardless of what method is used, the Panel recommends that the derivation of the TRVs is explicitly described and documented, and the specific TRVs selected fully justified to enhance the confidence that the TRVs that are selected are appropriate and protective. This is critical, since the selection of TRVs directly affects the risk assessment conclusions. A more useful approach would be to use a range of plausible TRV values for each receptor of concern, enhancing the utility of the results to the multiple-weight-of-evidence approach for risk management.

5.2 Other Recommendations

1. The Panel recommends an explanatory model-based approach to data evaluation over the calculation, and re-calculation, of uncertain hazard quotients up an ecological pathway tree. See Appendix A for an explanation of the explanatory model-based approach.

2. The Panel recommends analyzing the avian reproduction data using the Mayfield method (or similar approach; Mayfield, 1975; Johnson, 1979). This would account for differences in time of observation relative to nest initiation and other similar factors. It is standard practice in most avian productivity studies.

3. The Panel thanks MSU for providing additional information and data to supplement their publications. Because of the parsimonious nature of journal publications, it generally is not possible to incorporate the level of detail that is required for applying field-collected

data to a CERCLA risk assessment. Therefore, the Panel recommends that, in order to make the best use of the MSU studies in the KRSS risk assessment, a full data report be prepared. The report should include, for example, a synopsis of the Problem Formulation and Conceptual Model under which the data were gathered, referencing MSU's Work Plan and SOPs actually used as appropriate. Reasons for deviating from the stated objectives can be explained (for example, insufficient numbers of target species) and additional data beyond that presented in the publications can be presented. This would be particularly useful when applying the data to risk questions or approaches formulated after data collection or when merging MSU's data set with that of other data sources (e.g., BBL or CDM). For example, maps of the spatial relationships of the data, as shown at the September 2008 Panel Meeting, are of particular importance but not available from the publications. Similarly, descriptive data presentations such as scatter graphs or other graphical representations would present the data in ways that support in depth understanding and wider application. Log books, data tables, and other such details could be included as electronic appendices to the data report.

4. Use the MSU data in area-specific risk assessments — Given the knowledge of differences in habitat characteristics among impoundments, the MSU studies could provide data to assist development of area-specific conceptual models and exposure assessments.

5. Recognize the reality of floodplain dynamics and lack of independence of aquatic and terrestrial exposure routes — By demonstrating the significant linkages between aquatic and terrestrial food chains, the MSU studies may show that remedial actions to reduce exposures to aquatic biota would also reduce exposures to "terrestrial" avian species that feed at least in part on emergent aquatic insects.

6. Develop a more comprehensive conceptual model of the pathways of exposures to the various endpoints of concern. For instance, a comprehensive conceptual model might include the mink as one of the terrestrial receptors of concern. This species feeds on both aquatic and terrestrial animals and in the winter may rely primarily on terrestrial animals. Mink are known to be sensitive to the effects of PCBs. MSU studies have developed data on levels of PCBs in mink livers as an indication of exposure. Other opportunities presented by a comprehensive conceptual model include more clearly addressing the critical risk pathway identified by the BERA (via earthworms) and more clearly addressing the ecological risks in an altered future environment.

7. Perform quantitative uncertainty and sensitivity analyses on all models that utilize the MSU data, so that decision-makers are more fully informed about uncertainties inherent in these models and their parameterizations.

6.0 References

ARCADIS. 2008. *Baseline Ecological Risk Assessment Peer Review Scope of Work for the Allied Paper, Inc. /Portage Creek/Kalamazoo River Superfund Site*. March 2008.

Barnthouse, L. W., and G. W. Suter II. 1996. *Guide for developing data quality objectives for ecological risk assessments at DOE Oak Ridge Operations facilities*. ES/ER/TM-185/R1. Environmental Restoration Risk Assessment Program, Oak Ridge National Laboratory, Oak Ridge, TN.

Best, D.A., W.W. Bowerman, IV, T.J. Kubiak, S.R. Winterstein, S. Postupalsky. M.C. Shieldcastle, and J.P. Giesy. Jr. 1994. Reproductive Impairment of bald eagles along the Great Lakes shorelines of Michigan and Ohio. Pp. 697-702 in *Raptor Conservation Today, Proceedings of the IV World Conference on Birds of Prey and Owls*. B.-U. Meyburg and R.D. Chancellor, eds. Robertsbridge, U.K.: Pica Press.

Blankenship, A.L., M.J. Zwiernik, K.K. Coady, D.P. Kay, J.L. Newsted, K. Strause, C. Park, P.W. Bradley, A.M. Neigh, S.D. Millsap, P.D. Jones, and J.P. Giesy. 2005. Differential accumulation of polychlorinated biphenyl congeners in the terrestrial food web of the Kalamazoo River Superfund site, Michigan. *Environmental Science and Technology* 39:5954-5963.

Brenner, F.J., and D. Karwoski. 1985. Vocalizations as an index of abundance for the great horned owl (*Bubo virginianus*). *Proceedings of the National Academies of Science* 59:129-130.

Burnham, K.P., and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. *Second Edition.* Springer, New York.

Calabrese EJ, Baldwin LA. 1993. *Performing Ecological Risk Assessments*. Boca Raton, FL: Lewis Publishers. 257 pp.

CDM, 2003. *Final (Revised) Baseline Ecological Risk Assessment – Allied Paper, Inc./Portage Creek/Kalamazoo River Superfund Site*. Prepared on behalf of the MDEQ Remediation and Redevelopment Division. April 2003.

Chow, Shein-Chung, and Jen-pei Liu. 2008. *Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition.* Chapman & Hall/CRC Biostatistics Series.

Dahlgren, R.B., Linder, R.L., and Carlson, C.W. 1972. Polychlorinated biphenyls: Their effects on penned pheasants. Environ Health Perspect 1:89-101.

Duke, G. E., J. G. Ciganek, and O. A. Evanson. 1973. Food consumption and energy, water, and nitrogen budgets in captive Great Horned Owls (*Bubo virginianus*). *Comparative Biochemistry and Physiology* 44A: 283–292.

Efroymson, R. A., G. W. Suter II, B. E. Sample, and D. S. Jones. 1997. *Preliminary Remediation Goals for Ecological Endpoints*. ES/ER/TM-162/R2, Oak Ridge National Laboratory, Oak Ridge, TN.

Etterson, M.A., L.R. Nagy, and T.R. Robinson. 2007. Partitioning risk among different causes of nest failure. *The Auk*. 124:432-443.

Frank, R.A. 1997. *Effects of dieldrin on great horned owls at the Rocky Mountain Arsenal National Wildlife Refuge*. M.S. Thesis. University of Wisconsin, Madison, WI.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd Edition. Chapman & Hall, New York, NY. 668 pp.

General Electric Company (GE). 2005. *Interim Media Protection Goals Proposal for Housatonic River, Rest of River*. Can be downloaded from:http://www.epa.gov/NE/ge/thesite/restofriver-reports

Giesy, J., and K. Kannan. 1998. Dioxin-like and non-dioxin-like toxic effects of polychlorinated biphenyls (PCBs): Implications for risk assessment. *Critical Reviews in Toxicology* 28:511-569.

Giesy, J.P., J.P. Ludwig, and D.E. Tillitt. 1994a. Deformities in birds of the Great Lakes region: Assessing causality. *Environmental Science and Technology* 28:128A-135A.

Giesy, J.P., J.P. Ludwig, and D.E. Tillitt. 1994b. Dioxins, dibenzofurans, PCBs and colonial, fish-eating water birds. Pages 249-307 in A. Schecter, ed. *Dioxins and Health*. Plenum Press, New York.

Giesy, J. and M. Zwiernik. 2008. Ecological consequences of PCBs in the exposed sediments of formerly impounded areas of the Kalamazoo River. Presentation to the Kalamazoo River Ecological Risk Studies Peer Review Panel, 13 May 2008.

Hayward, J.L., J.G. Galusha, and G. Frias. 1993. Analysis of great horned owl pellets with rhinoceros aukelt remains. *The Auk* 110:133-5.

Holt, J.B., Jr. 1996. A banding study of Cincinnati area great horned owls. *J. Raptor Research* 30:194-197.

Houston, C.S. 1987. Nearly synchronous cycles of the great horned owl and snowshoe hare in Saskatchewa. Pp. 56-58 in R.W. Nero, R.J. Clark, R.J. Knapton, and R.H. Hamre (eds). *Biology and conservation of northern forest owls: symposium proceedings.* Gen. Tech. Rep. RM-142, U.S. Dept. Agric., Forest Service, Fort Collins, CO.

Houston, C.S., and C.M. Francis. 1995. Survival of great horned owls in relation to the snowshoe hare cycle. *The Auk* 112:44-59.

Houston, C.S., D. G. Smith, and C. Rhoner. 1998. Great horned owl. Pp. 1-27 in A. Pool and F. Gill (eds). *Birds of North America No. 372*. The Birds of North America, Inc., Philadelphia, PA

IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate Change 2007*. Four Volumes: a) *Synthesis Report.* b) *Working Group I Report, The Physical Science Basis.* c) *Working Group II Report, Impacts, Adaptation, and Vulnerability*. d) *Working Group III Report, Mitigation of Climate Change.* Intergovernmental Panel on Climate Change, United Nations Environmental Programme and World Meteorological Organization. Available at: http://www.ipcc.ch/index.htm.

Johnson, D.H. 1979. Estimating nest success: the Mayfield method and an alternative. *The Auk* 96(4): 651-661.

Keyser, A.J. J.T. Keyser, D.E.L. Promislow. 2004. Life-history variation and demography in western bluebirds (*Sialia mexicana*) in Oregon. *The Auk* 121:118-133.

Ludwig, J.P., H. Kurita-Matsuba, H.J. Auman, M.E. Ludwig, C.L. Summer, J.P. Giesy, D.E. Tillitt, and P.D. Jones. 1996. Deformities, PCBs, and TCDD-equivalents in double-crested cormorants (*Phalacrocorax auritus*) and Caspian terns (*Hydroprogne caspia*) of the upper Great Lakes 1986-1991: Testing a cause-effect hypothesis. *Journal of Great Lakes Research* 22:172-197.

Luttik, R., P. Mineau and W. Roelofs. 2005. A review of interspecies toxicity extrapolation in birds and mammals and a proposal for long-term toxicity data. *Ecotoxicology*. 14: 817-832.

Mayfield, H.F. 1975. Suggestions for calculating nest success. *The Wilson Bulletin* Vol. 87, No. 4 pp. 456-467.

Michigan Department of Natural Resources. 2002. Hunting Wildlife Habitat. Michigan's Wildlife Action Plan Terrestrial Systems: Southern Lower Peninsula. http://www.michigandnr.com/publications/pdfs/HuntingWildlifeHabitat/WCS/LF/SLP_Mesic_hardwood.pdf

Michigan State University Extension (MSUE). 2004. Michigan Natural Features Inventory. . http://web4.msue.msu.edu/mnfi/abstracts/ecology/Mesic_southern_forest.pdf

Mora, M.A., H.J. Auman, J.P. Ludwig, J.P. Giesy, D.A. Verbrugge, and M.E. Ludwig. 1993. Polychlorinated biphenyls and chlorinated insecticides in plasma of Caspian terns: Relationships with age, productivity, and colony site tenacity in the Great Lakes. *Archives of Environmental Contamination and Toxicology* 24:320-331.

MSU, 2001, Draft Interim Ecological Risk Assessment of Former Impoundment Soils

MSU. 2008. Overview of Studies Conducted by Michigan State University presented to the Review Panel at the meeting held at Brook Lodge, May 13 and 14, 2008

Neigh, A.M., M.J. Zwiernik, A.L. Blankenship, P.W. Bradley, D.P. Kay, M.A. MacCarroll, C.S. Park, P.D. Jones, S.D. Millsap, J.W. Newsted, and J.P. Giesy. 2006a. Exposure and multiple lines of evidence assessment of risk for PCBs found in the diets of passerine birds at the Kalamazoo River Superfund site, Michigan. *Human and Ecological Risk Assessment* 12:924-946.

Neigh, A.M., M.J. Zwiernik, P.W. Bradley, D.P. Kay, P.D. Jones, R.R. Holem, A.L. Blankenship, K.D. Strause, J.L. Newsted and J.P. Giesy. 2006b. Accumulation of Polychlorinated Biphenyls (PCBs) from floodplain soils by passerine birds. *Environmental Toxicology and Chemistry* 25: 1503-1511.

Neigh, A.M., M.J. Zwiernik, C.A. Joldersma, A.L. Blankenship, K.D. Strause, S.D. Millsap, J.L. Newsted, and J.P. Giesy. 2007. Reproductive success of passerines exposed to polychlorinated biphenyls through the terrestrial food web of the Kalamazoo River. *Ecotoxicology and Environmental Safety* 66:107-118.

Rhoner, C., and F.I. Doyle. 1992. Methods of locating great horned owl nests in the boreal forest. *Journal of Raptor Research* 26:33-35.

Sample, B. E., and C. A. Arenal. 1999. Allometric models for interspecies extrapolation of wildlife toxicity data. *Bulletin of Environmental Contamination and Toxicology* 62: 653-663.

Sample BE, Opresko DM, Sutter II GW. 1996. *Toxicological Benchmarks for Wildlife: 1996 Revision*. Risk Assessment Program, Health Sciences Research Division, US Department of Energy Oak Ridge National Laboratory, Oak Ridge, TN 37831.

Stephen, Charles E., Donald I. Mount, David J. Hansen, John R. Gentile, Gary A. Chapman, and William A. Brungs. 1985. *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*. PB85-227049. USEPA, Office of Research & Development, Environmental Research Laboratories, Duluth MN, Narragansett RI, Corvallis OR. National Technical Information Service, Springfield VA.

Stratus Consulting, Inc. 2005. *Stage I Assessment Report, Volume 1 — Injury Assessment: Kalamazoo River Environment Final Report.* Michigan Department of Environmental Quality, Michigan Attorney General, U.S. Fish and Wildlife Service, and National Oceanic and Atmospheric Administration.

Strause, K.D., M.J. Zwiernick, S.H. Im, P.W. Bradley, P.P. Moseley, D.P. Kay, C.S. Park, P.D. Jones, A.L. Blankenship, J.L. Newsted, and J.P. Giesy. 2007a. Risk assessment of great horned owls (*Bubo virginianus*) exposed to polychlorinated biphenyls and DDT along the Kalamazoo River, Michigan, USA. *Environmental Toxicology and Chemistry* 26: 1386-1398.

Strause, K.D., M.J. Zwiernick, S.H. Im, J.L. Newsted, D.P. Kay, P.W. Bradley, A.L. Blankenship, L.L. Williams, and J.P. Giesy. 2007b. Plasma to egg conversion factor for evaluating polychlorinated biphenyl and DDT exposure in great horned owls and bald eagles. *Environmental Toxicology and Chemistry* 26: 1399-1409.

Strause, K.D., M.J. Zwiernik , J.L. Newsted, A.M. Neigh, S.D. Millsap, C.S. Park, P.P. Moseley, D.P. Kay, P.W. Bradley, P.D. Jones, A. L. Blankenship, J.G. Sikarskie, and J.P. Giesy. 2008. Risk assessment methodologies for exposure of great horned owls (*Bubo virginianus*) to PCBs on the Kalamazoo River, Michigan. *Integrated Environmental Assessment and Management* 4:24-40.

US Environmental Protection Agency (USEPA). 1992. *Framework for Ecological Risk Assessment*. Washington (DC), USA: US Environmental Protection Agency Risk Assessment Forum. EPA/630/R-92/001.

US Environmental Protection Agency (USEPA) 1993. *Wildlife Exposure Factors Handbook*. USEPA Office of Research and Development. EPA/600/R-93/187)

U.S. Environmental Protection Agency (USEPA) 1994. *Guidance for the Data Quality Objectives Process*. EPA-600/R-96/055. Office of Research and Development, USEPA.

US Environmental Protection Agency (USEPA) 1997. *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments.* Interim Final. U.S. Environmental Protection Agency, Environmental Response Team (Edison, NJ). June 5, 1997.

US Environmental Protection Agency (USEPA). 1998. *Guidelines for Ecological Risk Assessment*. Washington (DC), USA: US Environmental Protection Agency. EPA/630/R-95/002F.

US Environmental Protection Agency (USEPA) 1999. *Screening Level Ecological Risk Assessment Protocol. Appendix E: Toxicity Reference Values*. USEPA Office of Solid Waste. Washington, DC.

US Environmental Protection Agency (USEPA) 2004. *Ecological Risk Assessment for General Electric (GE)/Housatonic River Site, Rest of River*. Can be downloaded from: http://www.epa.gov/NE/ge/thesite/restofriver-reports

US Environmental Protection Agency (USEPA) 2005. *Guidance for Developing Ecological Soil Screening Levels.* OSWER Directive 9285.7-55. USEPA Office of Solid Waste and Emergency Response. Washington, DC.

US Environmental Protection Agency (USEPA) 2007. *Ecological Soil Screening Levels for Polycyclic Aromatic Hydrocarbons*. Interim Final OSWER Directive 9285.7-78. USEPA Office of Solid Waste and Emergency Response. Washington, DC.

Warren-Hicks, W. J. 1999. Formal Methods for Risk-Based Decision-Making. *Human and Ecological Risk Assessment* 5(2):225-229.

Warren-Hicks, W. J., D. Moore. eds. 1998. *Uncertainty Analysis in Ecological Risk Assessment: Pellston '95*. SETAC Press, Florida.

Zwiernik, M.J., K.D. Strause, D.P. Kay, C.S. Park, A.L. Blankenship, and J.P. Giesy. 2007. Site-specific assessments of environmental risk and natural resource damage based on great horned owls. *Human and Ecological Risk Assessment* 13:966-985.

7.0 Appendix A - Suggested Statistical Analyses and Quality Assurance Evaluations

Before results between the BERA and MSU studies can be compared and contrasted, a basic assumption concerning the comparability of the data used in each study is required. The MSU study and BERA reach different conclusions. However, the reason for these differences is not readily apparent, but could be associated with such factors as: (1) data collection methods; (2) geographical differences in sampling locations; (3) differences in analytical chemistry methods; (4) differences in mathematical and statistical methods; (4) changes in biota or PCB concentrations over time; or (5) selection of toxicity threshold values. The Panel strongly suggests that a rigorous comparison be generated that illustrates the basic comparability of data collected in each of the studies. Below is a brief listing of a few of the many issues that should be addressed and presented to the reader in an understandable format:

- With the exception of Table 1-1 of the MSU Summary Document, a direct easy-to-read and understandable comparison of the data characteristics associated with each study is unavailable to the reader. The Panel strongly encourages such a table(s) (or figure) be developed. At a minimum, the comparable information should include sample location, number of samples, type of data collected, collection dates, etc. associated with each study.

- In keeping with the first point, the Panel suggests a similar one-to-one comparison table be created for the data analysis methods employed by each study. At a minimum, the reader should be able to easily juxtapose such information as TRVs used in the dietary and tissue hazard quotients for each endpoint, equations for back calculating cleanup values, average PCB concentrations (or TEQs) used in the numerator of the hazard quotients sorted by location, time, endpoint, trophic transfer factors used to calculate bioaccumulation and clean-up values, etc. Again, the reader should be provided ready access to information that will allow further investigation on how or why the two studies reach such different conclusions.

- Neither report addresses the comparability of the most basic data element, concentration of PCBs in various media. If indeed, a soil sample collected by MSU results in a different concentration than a replicate sample collected by EPA, then a comparison of statistical outputs using the two independent data sets is compromised. Therefore, the reader must be convinced that the studies are actually evaluating the same concentration information (including PCB concentrations, TEQs, etc).

- The effect of time is not addressed by either study. It seems that among the two studies, data are collected over many years. The investigators seem to have ignored the role that time effects can play on the comparability of the two data sets and associated findings. A rigorous evaluation of the effect of time on the discrepancy in the study findings should be implemented.

Adequate sample sizes: The Panel has criticized the MSU studies for small sample sizes in estimation of many parameters. Advance planning for adequate sample sizes is best conducted by guessing the width of future confidence intervals on parameters, based on guesses of variation in future data. Betting on the future, based on past data, has its obvious drawbacks.

After a study is completed, the question often arises, was the sample size adequate? To answer this question is not easy. A good criterion is “Does the sample size pass the laugh test based on prior experience and scientific knowledge?” A more “scientifically” acceptable approach is to emphasize point estimation of parameters with confidence intervals or other measures of precision rather than tests of null hypotheses (i.e., hypotheses that two or more parameters are equal). If the widths of confidence intervals on important parameters are narrow enough to allow clear conclusions to be made, then the sample sizes were adequate.

Pseudo-replication: Three basic forms of pseudo-replication were present in the MSU studies. One source is because study units (e.g., nest boxes, sampling grids for collection of soil, etc.) were not selected by some probabilistic procedure (e.g., simple or systematic sampling of points for location of nest boxes) from the entire study. This form of pseudo-replication implies that conclusions are not necessarily valid for the entire study areas. Even if conclusions are limited to the “sub-regions” studied within the study areas, a second source of pseudo-replication exists, namely that study units were potentially dependent in the sense that knowledge of the measurement of a parameter on one unit gives information on other nearby units. The effective sample sizes are less than the stated values. A third source of pseudo-replications came from pooling of sample data over time from the same study units, e.g., bird reproduction data from the same nest box. The end results of pseudo-replication in the MSU studies are: (1) the study design does not guarantee that the conclusions apply to the entire study areas, and (2) the “true” precision of point estimates and “p-values” of tests of null hypotheses are greater than the stated values by unknown amounts.

Measures of Central Tendency. The geometric mean underestimates the central tendency of lognormal distributed variables. Other measures of central tendency include the median and arithmetic mean of a set of measurements. See the attached Figure 2.1, for the location of the geometric mean, median, and mean in an example sample of size 30 from a lognormal distribution. The Panel suggests that the median or arithmetic mean of original data (not log-transformed data) be used for all analyses calling for measures of central tendency of exposure data. The arithmetic mean will provide the most conservative (largest) estimate of central tendency of concentration of PCBs in food items because extremely large concentrations of PCBs are encountered with small probabilities (assuming the values follow a log-normal distribution with its positively skewed distribution). The median or arithmetic mean should be used with this explicit understanding (e.g., they provide the "middle" and most conservative measures of central tendency of concentration).

Recommended Statistical Analyses: For the simpler parameters, analyses should be based on point estimation of parameters with measures of precision such as standard errors and confidence intervals. Confidence intervals should be computed for parameters and the difference or ratio of parameters, then plotted graphically so that potentially important biological differences can be seen. Alternatively, box-and-whisker plots of data collected from two different sites or times can be displayed in graphical form side-by-side. Unfortunately, conclusions based on confidence intervals and other methods will continue to be limited by the small sample sizes and pseudo-replication in the MSU studies. Consider Table 2-1 in the Overview of Studies Conducted by Michigan State University presented to the Review Panel at the meeting held at Brook Lodge, May 13 and 14, 2008, and Neigh et al. (2007, number 4 in the papers provided to the review team). See bottom of page 110 and page 111 in Neigh et al. (2007). For example, small sample sizes within a year or acceptance of a null hypothesis of 'no difference among years' are not justification for combining reproductive data of eastern bluebirds or house wrens among all years. The decision to pool data from different sources is a subjective decision, not a statistical inference. Confidence intervals could be computed for each year of the reproductive parameters and plotted next to each other on the same figure to provide useful information concerning the differences and degree of variation within each year. Alternatively, box-and-whisker plots adjacent to each other will provide essentially the same information. Generally, models are fitted to data from various sources in time and space. That is, in addition to the plots over time and sites, multiple regression models could be fitted to explore the relationships between reproductive parameters and predictor (independent) variables such as: year, early versus late nests, sites within Fort Custer and within Trowbridge, Fort Custer versus Trowbridge, etc. Tests of hypotheses and measures of precision associated with the models are subject to question because of the pseudo-replication in these studies. Granted that the statistical inferences are limited, useful models may be obtained. ANOVA was conducted using linear models in some cases, however only for the inappropriate purpose of testing null hypotheses. Models for prediction of observed effects on parameters, as functions of covariates measured on the study sites and times, should be developed using variations of the *Akaike's information criterion* (AIC) for selection among competing models (see, e.g., Burnham and Anderson 2002).

In short, tests of null hypotheses should probably not be used in any of the MSU statistical analyses, a realization that is beginning to take root in many disciplines of science. If tests of hypotheses are to be used in evaluating impacts of PCBs or other toxicants, then they should be stated in terms of 'tests of bioequivalence' (see for example, Chow and Liu 2008). A test of bioequivalence would hypothesize that there is an important effect of PCBs unless the data are sufficient to prove that, for example, reference and impoundment sites are "bioequivalent." Essentially, the concentrations of PCBs in the impoundments are assumed "guilty until proven innocent" by the data. Small sample sizes are acceptable from a regulatory agencies' point of view when tests of bioequivalence are used, because small sample sizes would likely fail to prove that the PCBs are "innocent."

Conclusions based on point estimates with measures of precision (e.g., confidence intervals) or tests of bioequivalence will be subject to some of the same limitations as discussed above because pseudo-replication exists in the MSU data, i.e., the study design does not guarantee that the conclusions from confidence intervals or tests of bioequivalence apply to the entire study areas. However, small sample sizes yield wide confidence intervals, and small sample sizes will tend to not be able to reject the assumption that there are important effects of PCBs. Results will be more informative than the tests of null hypotheses conducted by MSU.

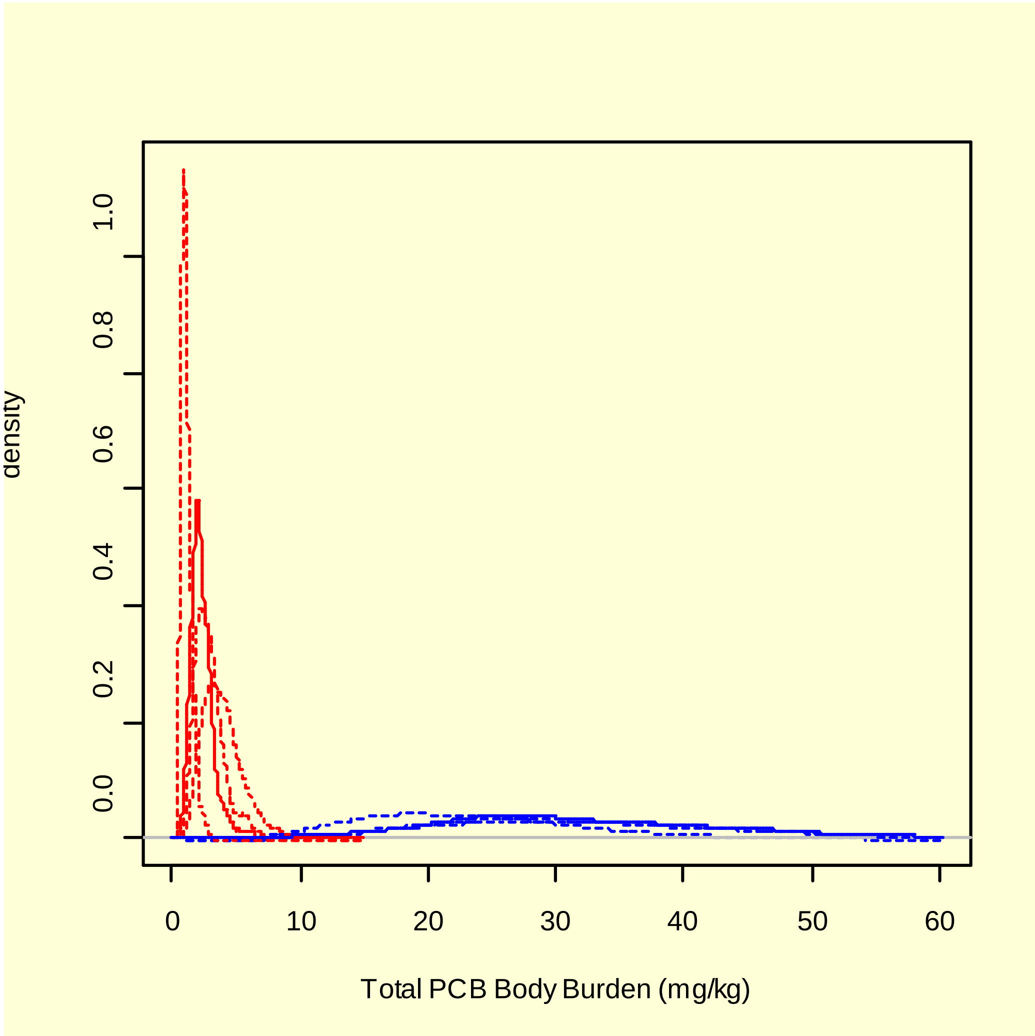
Development of Explanatory Models: The MSU investigators did not take advantage of model-based analyses and the relatively rich data set they collected. Condensing the data to simple hazard quotients is under-utilization of a valuable and rich dataset. An advantage of the MSU data relative to the BERA data is that MSU emphasized data on the response endpoints of interest and site-specific measures of PCB uptake. Granted there are problems with pseudo-replication and small sample sizes, MSU is afforded the ability to explore relationships among the data variables, examine the distribution of these variables over time and space, and use this information to draw valuable model based inferences on the PCB exposure potential and the relationship of PCB concentrations to effects.

As an example, the following model was used in the analysis of PCB effects on the Housatonic River (http://www.epa.gov/region1/ge/thesite/restofriver-reports.html#Eco).

$$T_{BodyBurden} = (MT + \sum_{i=0}^{i=14} (FIR_i \ x \ C_{diet} \ x \ 1day) \ / \ BW_{14}$$

$$FIR(kg / kg \ bw / day) = \frac{\alpha \ BW^{\beta}}{K}$$

The Panel understands that MSU collected data on PCB body or tissue burden (T), egg concentration (MT), concentration in diet (C), and body weight (BW). These types of models should be used to explore relationships among the endpoints of interest (e.g., body or tissue burden among others) and co-variables (egg concentration, diet, etc.) in an effort to generate additional insights that are achievable beyond hypothesis testing or simple hazard quotients. For example, the Panel understands that MSU has data for calibrating the above model at several locations within Trowbridge and in the reference area. If that is the case, then information like that generated at the Housatonic River (below), through the model, can be used as an alternative to hypothesis testing to infer the magnitude, uncertainty, and geographic differences in the endpoints of interest.



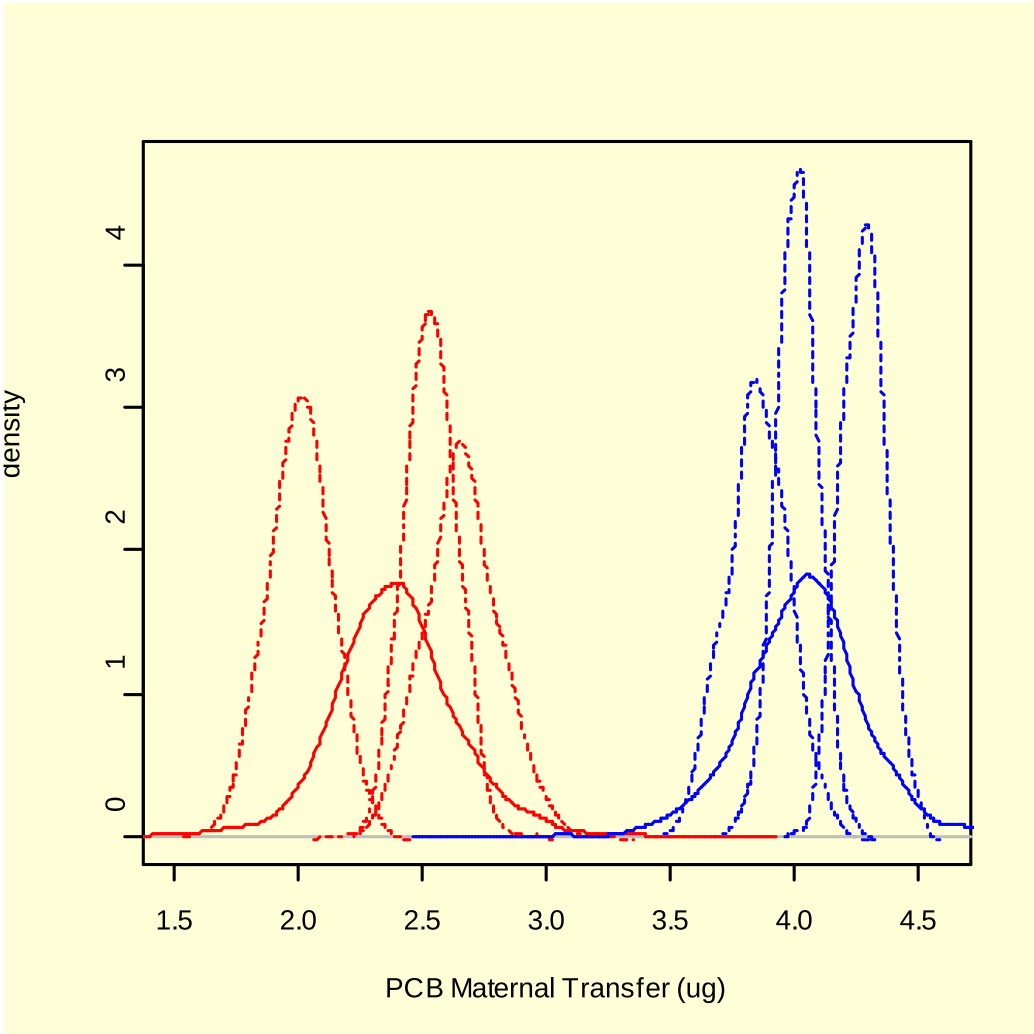
In the above figure, the red dotted lines represent the distribution of PCB body burden at each of three reference sites, and the solid red line represents integration across all reference sites. Notice that the distributions are tall and thin with similar centers, indicating little uncertainty among and within the reference sites. In contrast, the blue lines indicate the impact sites that show a great deal of uncertainty about the center of the distribution, but are easily seen to be different on average than the reference sites. Such probabilistic/graphical analysis of the data provides an insightful way of generating inferences from the data without the use of hypothesis testing.

Co-variables like maternal transfer (egg concentrations) can be treated in the same way (see above figure). And measures of "so what" (i.e., the effects endpoints) can be overlain on such exposure distributions to aid in decision-making. In fact, if distributions of effects (not employed by either EPA or MSU) are available, then an integration of the exposure and effects distributions may prove useful. The integration can be preformed graphically or numerically.

The Panel believes that other mathematical/probabilistic model forms can also be devised and implemented for the variety of data types and variables represented by the MSU data sets. Also note that such models and graphical outputs can be used as an aid toward resolving extrapolation issues. Suppose one wishes to extrapolate the results found at one site to another; for example, suppose there are measurements of the model response variable at Trowbridge but not downstream at Otsego. However, if co-variable measurements are available at Otsego, then statistical methods exist that will allow the extrapolation of the Trowbridge information, through the model, to estimate the expected body (or tissue) burdens at Otsego (see Gelman et al. 2004).

In any case, because the MSU investigators generated substantially different findings than BERA, a more extensive analysis of the data should be conducted by MSU. Given that the MSU data information content is richer than that collected by USEPA, the Panel believes that additional data analysis approaches should be used to expand the insights available from the information.

Formal Uncertainty and Sensitivity Analyses for Characterizing Risk: Neither BERA nor MSU implemented any sort of formal uncertainty analysis in their respective approaches to establishing risk. Again, given the differences in risk characterization among the studies, the Panel strongly believes that both USEPA and MSU should conduct formal uncertainty analyses. MSU, for example, could generate exposure and effects distributions using probabilistic techniques (rather than the simple hazard quotients). Similarly, simple sensitivity analyses of hazard ratios formed with differing estimates of exposure and effect could be implemented and graphically presented.



At the end of the day, investigators associated with the BERA and MSU studies should examine the degree of overlap among the competing data sets and analytical outputs and decide whether or not the findings are significantly different within the bounds of the available information content represented by the collected metrics.

A comprehensive listing of mathematical approaches for conducting these analyses is not presented here, but can be found in Warren-Hicks and Moore (1998) and Warren-Hicks (1999).

Pooling Data Across Studies: The MSU summary states:

1. The information was collected with the primary goal of addressing uncertainty in the Baseline ERA (CDM 2003).

2. Our intent was to develop multiple, site-specific, independent lines of evidence to supplement those evaluated in the Baseline ERA …

Given the above goals, particularly the notion that the MSU data were collected to supplement the BERA lines-of-evidence, the Panel recommends that BERA and MSU should attempt to merge the data sets, or at least provide a one-to-one evaluation of the measured metrics in each study. There is a number of advantages to pooling the information across studies, including creating a longer time-series of information, increased information content, increased geographical scale, and an increased ability to draw inferences based on the data information content.

Several methods are available for pooling data, including: (1) simple concatenation of data sets using expert judgment to identify those cases where the data cannot be pooled based on scientific reasoning; (2) formal methods for pooling based on underlying probability distributions; and (3) updating approaches when the data are time-dependent (see Gelman et al. 2004). However, there are many measures, such as tissue concentrations and body burdens, that BERA generated from literature values and MSU measured in the field. In these cases, the field measurements are generally preferred and datasets should not be combined.

The analyses described above should only be implemented after an evaluation of the ability to pool data from the studies is completed.

Time-Series Analyses: Data used in the MSU analyses were generally collected from 2001 – 2003. Little or no information is provided in the BERA on the time spans over which data were collected and compiled for the various analyses employed by BERA (see, for example Table C-1 and notice that data characteristics including data collection times are not provided).

The investigators need to demonstrate that the data-collection time is not a factor underlying the discrepancy in risk characterization results among the studies. Time-series plots, using data from both studies (see above comment), should be generated. Hypothesis testing should not be used as a basis for pooling data over time (see above comment). If specific metrics are shown to have time trends or cycles, the effect of this observations on the risk characterization results must be described, and specific mathematically defensible methods for formally incorporating a time component into the risk analyses must be implemented.

Tiered Risk Assessment: Both MSU and BERA have effectively employed simple statistical and data analytic approaches for evaluating the data, typically those employed during the early tiers of a formal risk analysis. Given the discrepancy in risk characterization among the studies, more advanced statistical and risk characterization techniques (like those described above) are warranted. In particular, the Panel encourages MSU and USEPA to reduce the dependence of the risk decisions on hazard quotients, and implement techniques that make full use of the available information. In particular, uncertainty analyses, time-series evaluations, descriptive graphical analyses, and explanatory models should be used to further evaluate the data and provide insights into the differing risk decisions generated by the BERA and MSU.